UNIVERSITÉ DE MONTRÉAL

OPTIMISATION SOUS CONTRAINTES D'UN MODÈLE HYDROLOGIQUE POUR UNE REPRÉSENTATION DE LA PHYSIQUE DES PROCESSUS

DOMINIQUE CARTIER
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(MATHÉMATIQUES APPLIQUÉES)
OCTOBRE 2012

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

OPTIMISATION SOUS CONTRAINTES D'UN MODÈLE HYDROLOGIQUE POUR UNE REPRÉSENTATION DE LA PHYSIQUE DES PROCESSUS

présenté par : CARTIER Dominique
en vue de l'obtention du diplôme de : Maîtrise ès Sciences Appliquées
a été dûment accepté par le jury d'examen constitué de :

M. SAVARD Gilles, Ph.D., président
M. AUDET Charles, Ph.D., membre et directeur de recherche
M. LE DIGABEL Sébastien, Ph.D., membre et codirecteur de recherche
Mme SIMARD Marie-Claude, M.Ing., membre

À ma fille Maëly qui donne un sens à mes efforts,
à mon mari Nicolas pour ses encouragements,
à mes parents pour leur soutien et leurs petites attentions…

REMERCIEMENTS

Ce mémoire cache toute une équipe. Mon nom figure sur la couverture, mais c'est seulement parce que la page n'était pas assez grande pour accueillir tout ceux qui mériteraient de s'y trouver. Ainsi, je tiens premièment à souligner la grande disponibilité de mes directeurs de recherche, Charles Audet et Sébastien Le Digabel. Tout au long de mon périple, vous avez été présents pour répondre à mes interrogations et inquiétudes. Vous avez su me prodiguer des conseils avisés et réguliers durant mes longs mois de stage à l'Institut de recherche d'Hydro-Québec (IREQ) et m'avez aussi fourni les outils dont j'ai eu besoin au fil du temps. Vous m'avez généreusement fait profiter de bourses de recherche afin que tous mes efforts et toutes mes énergies soient concentrés sur ma recherche. Finalement, je vous suis grée de toutes les heures passées à la relecture de ce document. Vu les délais très courts, vous avez dû prioriser mon mémoire au détriment de vos autres projets et cela a été très apprécié.

Je tiens aussi à mettre en évidence que ces travaux n'auraient pas été possibles sans la collaboration du consortium Ouranos, qui a gracieusement généré et fourni les données du MRCC.

À l'Institut de recherche d'Hydro-Québec, c'est toute une unité que j'aimerais remercier pour m'avoir fait sentir à ma place dès le début. Merci donc, unité Métallurgie, Mécanique et Hydroéolien, et spécialement à Dominique Tapsoba, pour m'avoir si bien intégrée et pour avoir pris à cœur mes travaux, mais aussi ma vie familiale et scolaire. Une mention aussi à James Merleau pour ses précieux conseils statistiques, à Jean Leduc et Manon Vincent, pour avoir permis et géré mes multiples retours à l'IREQ et à René Roy, chargé de projet « Climatologie », pour l'opportunité de stage.

Un merci tout particulier à la meilleure superviseure de stage que le monde ait portée. Marie Minville, non seulement tu m'as accordé beaucoup de ton temps, mais en plus tu as été très accueillante. J'ai eu beaucoup de plaisir à travailler avec toi et j'espère avoir bientôt l'occasion de remettre ça. Je voudrais aussi profiter de l'occasion pour souligner la pertinence de ton idée de calage parallèle. Non seulement as-tu vu juste, mais tes intuitions tout au long de ce projet ont porté fruits et ont permis de pousser la réflexion à un autre niveau.

Merci à Catherine Guay pour ta bonne humeur, tes connaissances et ton point de vue extérieur qui nous a permis de voir le problème sous un autre angle. Merci aussi de m'avoir

donné l'opportunité de participer au projet d'envergure qu'est le projet $(CQ)^2$.

Finalement, je me dois de mentionner la contribution de Louis-Alexandre Leclaire à ce projet. Non seulement as-tu développé EAUPTIM, ce complexe outil d'optimisation et de simulation, impressionnant de par sa taille mais aussi de par toutes ses subtilités. Tu as aussi offert un service après-vente exemplaire ! Tu m'as formée à l'utilisation de tous les outils informatiques dont j'ai eu besoin, tu m'as assistée dans mes tests et tu as répondu à mes appels de détresse. Ton aide dans ce projet a été essentiel et je tiens à t'exprimer ici ma gratitude.

Sur une note plus personnelle, j'aimerais remercier du fond du cœur mes parents pour leur soutien tout au long de mon périple académique, et particulièrement durant les dernières semaines de travail acharné précédant le dépôt de mon mémoire. Les sorties surprises avec Maëly afin que je puisse travailler dans le calme et les trucs de grand-mère afin d'augmenter ma concentration ont été plus qu'appréciés. Sans vous, je ne serais probablement pas parvenue à boucler la boucle dans les temps et je vous en suis très reconnaissante.

Dans le même ordre d'idées, je tiens à remercier haut et fort mon mari Nicolas, qui, durant les dernières années, a dû mettre les bouchées doubles au travail pour me permettre d'être sur les bancs d'école. Merci de prendre soin de moi et de notre cocotte, merci d'être là dans les moments difficiles et de me comprendre dans les périodes plus stressantes. De tout cela et de bien d'autres choses encore, je te suis très reconnaissante.

La dernière et non la moindre, merci à ma petite fille Maëly. Certaines personnes disaient que ce serait plus difficile de réussir avec un enfant, mais au contraire ! Grâce à toi, j'ai passé à travers chaque journée avec un sourire et tu m'as donné une raison de continuer quand les choses devenaient corsées. Tu mets du soleil dans ma vie et je t'aime très fort.

RÉSUMÉ

Les modèles hydrologiques sont utilisés sur une base quotidienne par nombre d'experts notamment pour dimensionner des ouvrages, planifier la production d'hydroélectricité ou limiter les dommages liés aux inondations. Ces simulations informatiques sont conçues afin de reproduire le cycle de l'eau par le biais d'équations afin d'arriver à une estimation fiable des débits à l'exutoire d'un bassin versant. Des variables d'optimisation sont présentes au sein de ces équations et la principale difficulté liée à l'utilisation de ces modèles consiste à trouver l'ensemble de valeurs produisant la série de débits la plus près de celle réellement observée. Cette opération se nomme calage ou étalonnage de modèles hydrologiques et offre une paramétrisation réalisée à partir de données historiques.

Dans un contexte de réchauffement planétaire, des entreprises comme Hydro-Québec se doivent de produire des projections futures des débits afin de mesurer l'impact des changements climatiques sur leurs activités. Une paramétrisation basée sur le passé soulève un doute quant à son applicabilité en période future, où les événements hydrométéorologiques pourraient être très différents. De plus, la formulation actuelle du problème d'optimisation ne permet pas d'encadrer les variables intermédiaires associées aux processus hydrologiques, comme la neige, l'évapotranspiration ou le débit de base. Ainsi, un ensemble de paramètres est fourni à la simulation et seule la sortie du modèle, donc les débits, est considérée. L'impact est une représentation des processus hydrologiques ne respectant pas les lois de la physique, ce qui renforce l'hypothèse selon laquelle la présente paramétrisation du modèle ne peut, en toute confiance, être transposée en contexte de changements climatiques. L'objectif visé par ce mémoire est donc, par le biais des outils d'optimisation, de présenter une méthode de calage orientée vers la physique des processus et offrant une paramétrisation robuste en contexte de changements climatiques ; notre hypothèse de recherche étant qu'un ensemble de paramètres offrant un respect de la physique dans la représentation des processus hydrologiques est préférable a une solution présentant une meilleure valeur de la fonction objectif, mais modélisant les processus en toute liberté.

Ainsi, deux stratégies de calage ont été mises de l'avant et sont comparées à la méthode largement utilisée dans la communauté scientifique, notamment chez Hydro-Québec. La première stratégie proposée vient, par le biais d'une contrainte, recentrer le bilan de l'évapotranspiration réelle sur la valeur offerte par l'Atlas du Canada, alors que la seconde, toujours via une contrainte, met à profit l'immense base de données qu'offrent les modèles climatiques en

contraignant la forme du cycle de l'évapotranspiration réelle. Cette approche par contraintes a été raffinée par l'utilisation d'un outil d'analyse de sensibilité des contraintes offert en complément de l'implémentation de MADS (NOMAD). Chaque contrainte fut ainsi traitée afin d'être relâchée ou resserrée selon la sensibilité de la fonction objectif et selon la confiance en la qualité des données observées, souvent entachées d'erreurs. Ce sont donc des milliers d'optimisations qui ont été réalisées par l'algorithme MADS afin d'arriver à une comparaison des différentes méthodes sur le plan de la juste représentation des hydrogrammes et des évapogrammes (ou cycle annuel moyen de l'évapotranspiration), ainsi que sur la variabilité de la paramétrisation et la robustesse des solutions.

De ces analyses, il est ressorti que le calage parallèle offre la meilleure performance pour plusieurs raisons. Une forte corrélation a été remarquée entre la qualité des solutions produites par l'optimisation et celles produites en simulant les débits d'un ensemble de données n'ayant pas servi à la calibration, dites de validation. Ainsi, il est possible de conclure à une solide robustesse pour une utilisation en contexte de changements climatiques. La meilleure solution obtenue par la stratégie de calage parallèle s'est aussi révélée capable de produire une meilleure représentation des débits en validation que l'ensemble des stratégies, le tout en modélisant le processus d'évapotranspiration de façon à respecter la physique.

L'application la plus notoire de la méthode développée dans ce mémoire est l'adoption de la stratégie de calage parallèle pour le projet $(CQ)^2$. Ce projet, collaboration d'Hydro-Québec, du consortium Ouranos sur la climatologie régionale et l'adaptation aux changements climatiques, du Centre hydrique du Québec (CEHQ) et de l'entreprise Rio Tinto Alcan, vise à qualifier et à quantifier les impacts des changements climatiques sur le régime hydrologique québécois. Pour y arriver, une paramétrisation unique a été optimisée à l'aide du calage parallèle pour chacun des 324 bassins versants recouvrant l'ensemble du territoire québécois. Au terme de ce projet, ce sont presque 160 000 scénarios hydroclimatiques qui auront été générés, soit 98 simulations climatiques employant 5 différentes méthodes de post-traitements déployées sur les 324 bassins versants. Cette base de données considérable n'a de valeur que si l'étalonnage du modèle est valide et c'est donc une grande marque de confiance des décideurs du projet envers les résultats obtenus dans cette recherche que d'avoir choisi notre méthode pour la calibration du modèle HSAMI.

Les contributions de ce mémoire à l'avancement des connaissances en hydrologie sont nombreuses. La plus flagrante est la présentation d'une méthode de calage robuste pouvant être transposée en contexte de changements climatiques et pouvant être étendue à plusieurs

autres processus hydrologiques pour une représentation encore plus orientée vers la physique. D'ailleurs, l'intégration de données complémentaires provenant de modèles climatiques dans le processus d'optimisation est proposée pour la première fois à la communauté scientifique. Le regard nouveau porté sur les processus d'optimisation a permis une amélioration non négligeable de la performance de l'algorithme d'optimisation par le biais de l'emploi approprié des options offertes. L'utilisation du module complémentaire d'analyse de sensibilité a aussi ouvert de nouvelles portes en offrant la possibilité de mettre à profit les connaissances des utilisateurs d'un modèle dans la formulation adéquate des contraintes. L'approche de calage par contraintes proposée dans cette recherche s'en est vue bonifiée et de nouvelles avenues de recherche en ont découlé. Finalement, pour le domaine de la recherche opérationnelle, ce projet a offert une belle opportunité d'application des outils de l'optimisation. Non seulement l'algorithme MADS est-il mieux compris par les chercheurs de l'IREQ, mais le module d'analyse de sensibilité s'est fait connaître et continuera d'être employé.

ABSTRACT

Hydrological models are used on a daily basis by many experts in particular to design dams, to plan hydropower production or reduce flood damages. These computer simulations are employed to model the water cycle through equations in order to achieve a reliable estimate of flows at the outlet of a watershed. Optimization variables are present in these equations and the main goal is to find the set of values producing the flow series closest to the one actually observed. This operation is called calibration of hydrological models and provides a parameterization using historical data.

In a context of global warming, companies such as Hydro-Québec must produce projections in order to measure the impact of climate changes on their activities. Parameterization based on the past hydrological series raises doubt as to its applicability in future periods, where hydrological events could be different. In addition, the current formulation of the optimization problem does not allow to monitor intermediate variables associated with hydrological processes, such as snow, evapotranspiration or base flow. Thus, a set of parameters is provided to the simulation and only the output of the model, the flow, is considered. The impact can be a representation of hydrological processes not necessarily complying with the physical laws, which reinforces the hypothesis that the manner in which models are calibrated cannot confidently be transposed in the context of climate changes. Based on optimization tools, the objective of this paper is to develop a calibration method oriented on physical processes and provide a robust parameterization in the context of climate changes; our research hypothesis is that a set of parameters modeling hydrological processes with respect to the physical representation leads to a better solution than one which relies only on a better value of the objective function, modeling hydrological processes with complete freedom.

While two calibration strategies are put forward and are compared to the method currently used at Hydro-Québec. The first strategy which is proposed focuses on the assessment of the actual evapotranspiration of the value offered by the Atlas of Canada through a constraint. The second uses the database offered a climate model by constraining the shape of the actual evapotranspiration cycle. This approach with constraints is refined using a sensitivity analysis tool available in the NOMAD package. Each constraint is thus treated to be relaxed or tightened depending on the sensitivity of the objective function and according to the confidence in the quality of the observed data, often marred by error. Thousands of optimization are performed by the MADS algorithm to compare different methods in terms of a good rep-

resentation of hydrographs and evapographs (or average annual cycle of evapotranspiration), as well as the variability of "optimal" sets of parameters and the robustness of these solutions.

From these analyses, the parallel calibration was found to offer the best performance on many levels. A strong correlation is found between the solutions produced by the optimization and those produced by simulating the flow of a set of data not used for calibration, called validation. Thus, it is possible to conclude to a stronger robustness in the context of climate changes. The best solution obtained by the parallel calibration strategy is also proved capable of producing a better representation of the validation flows than the other strategies, while modeling the evapotranspiration process in order to respect physical process.

The most remarkable application of the method developed in this thesis is the adoption of parallel calibration strategy for the project $(CQ)^2$. This project, which brings together with Hydro-Québec, the Ouranos consortium on regional climatology and climate changes adaptation, the Centre d'expertise hydrique du Québec (CEHQ) and Rio Tinto Alcan, aims to qualify and quantify the impacts of climate changes on the hydrological regime in Quebec. To attain this goal, a single parameterization has been optimized using parallel calibration for each of the 324 watersheds covering the entire province of Quebec. At the end of this project, almost 160 000 hydroclimatic scenarios have been generated, from which 98 climate simulations using five different methods of post-processing deployed on 324 watersheds. This considerable database only has value if the model calibration is valid and it is thus a great vote of confidence from the project decision makers towards the results obtained in this research to have chosen our method for the calibration of the HSAMI Model.

The contributions of this thesis to the advancement of knowledge in hydrology are numerous. The most obvious is the development of a robust calibration method which can be implemented in the context of climate changes and can also be extended to several other hydrological processes for better physical representation. In fact, it is the first time that data from a climate model are integrated in the calibration process. This new perspective on the optimization procedure results in a significant improvement of the performance of the optimization algorithm through the appropriate use of the offered options. The use of a complementary sensitivity analysis module also opens new doors in offering the possibility to better take advantage of the user knowledge of a model in formulating appropriate constraints. The calibration constraint approach proposed in this research is thus enhanced and new avenues of research result from it. Finally, for the operational research field, this project provides a great opportunity for the application of optimization tools.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ANNEXES

LISTE DES SIGLES, ABRÉVIATIONS ET NOTATIONS

Sigles et abréviations (hydrologie)	
ETR	Évapotranspiration réelle
HSAMI	Service hydrométéorologique apports modulés intermédiaires
IREQ	Institut de recherche d'Hydro-Québec
krigIREQ	Données météorologiques obtenues par krigeage à l'IREQ
MRC	Modèle régional du climat
NLWIS	Service national d'information sur les terres et les eaux (*National Land and Water Information Service*)
ns	Critère de Nash-Sutcliffe
P	Précipitations totales
Q	Débit
Qfc	Transpiration de la végétation
Qfcf	Évaporation de la pluie interceptée par la neige
Qfcl	Évaporation de la pluie interceptée par la végétation
Qfg	Évaporation du sol nu
Qfn	Sublimation et évaporation de l'eau contenue dans la neige
Qfs	Flux de vapeur d'eau vers l'atmosphère (évapotranspiration) en mm/j

Sigles et abréviations (optimisation)	
BiMADS	MADS pour les problèmes bi-objectif
CS	Recherche par coordonnées (*Coordinate Search*)
EB	Barrière extrême (*Extreme Barrier*)
F	Filtre (*Filter*)
GPS	Recherche par motifs généralisée (*Generalized Pattern Search*)
LT-MADS	MADS avec matrices triangulaires inférieures (*Lower Triangular MADS*)
MADS	Recherche directe sur treillis adaptatif (*Mesh Adaptative Direct Search*)
NOMAD	Optimisation non-linéaire avec MADS (*Nonlinear Optimization by Mesh Adaptative Direct Search*)
OrthoMADS	MADS déterministe avec directions orthogonales (*Orthogonal MADS*)
PB	Barrière progressive (*Progressive Barrier*)
PEB	Barrière progressive-extrême (*Progressive-to-Extreme Barrier*)

PSD-MADS	MADS avec décomposition parallèle de l'espace (*Parallel Space Decomposition with MADS*)
VNS	Recherche à voisinage variable (*Variable Neighborhood Search*)

Notations	
Δ_m^k	Paramètre de taille du treillis à l'itération k
Δ_p^k	Paramètre de taille de sonde à l'itération k
Ω	Ensemble des points réalisables
X	Sous-ensemble de $\mathbb{R}^n$ comprenant les points respectant les contraintes non relaxables et non quantifiables
$c_j(x)$	Membre de gauche des contraintes relaxables et quantifiables
D_k	Ensemble de directions généré à l'itération k
$f(x)$	Fonction à optimiser
m	Nombre de contraintes relaxables et quantifiables
n	Nombre de variables
Qfs_{moy}	Moyenne de la variable d'évapotranspiration du modèle climatique
$\overline{Q_o}$	Moyenne des débits observés
Q_o^t	Débit observé à chaque pas de temps
Q_s^t	Débit simulé à chaque pas de temps
x_0	Point de départ
x_k	Solution à l'itération k
y	Point d'essai à tester à l'itération k

CHAPITRE 1

INTRODUCTION

Les modèles hydrologiques sont utilisés à travers le monde par un nombre impressionnant d'experts. Que ce soit afin de prévoir les apports d'un bassin versant, de dimensionner certains ouvrages ou d'estimer les impacts des changements climatiques sur la ressource en eau, leurs usages sont très variés. Chez Hydro-Québec, une équipe de sept prévisionnistes utilise chaque matin le modèle hydrologique HSAMI afin de prévoir le débit à l'exutoire de 90 bassins versants faisant partie du parc hydroélectrique. Des prévisions pour la journée, mais aussi pour les 10 et 200 prochains jours sont fournies aux gestionnaires des barrages afin qu'ils puissent gérer efficacement l'eau et le parc de production (maintenance de l'équipement, centrales au fil de l'eau, usages multiples de l'eau, etc.).

L'utilisation de ces modèles en mode simulation est très simple et, pour les modèles plus simples dits conceptuels, ne consomme pas beaucoup de temps ni de puissance de calcul. Ainsi, des données météorologiques (et parfois physiographiques selon la complexité du modèle) sont fournies en intrants au modèle qui, par une série d'équations, tentera de reproduire le cycle de l'eau pour finalement arriver à une estimation des débits pour une période donnée. Le nœud du problème réside dans la calibration du modèle, aussi appelé calage, étalonnage ou optimisation, et signifiant la recherche d'un jeu de paramètres [1], présents dans les équations, permettant une série de débits simulés s'ajustant le mieux possible aux données observées pour un bassin versant donné.

1.1 Étalonnage de modèles hydrologiques

La petite histoire de l'optimisation des modèles hydrologiques remonte à l'an 1674, où le premier modèle hydrologique, celui de Perreault, a été élaboré (Chkir (1994)). Depuis lors et jusqu'à la fin du vingtième siècle, la calibration de ces modèles se faisait à la main. Ainsi, plusieurs jeux de paramètres étaient soumis au modèle et les simulations résultantes étaient comparées par des hydrologues afin de statuer sur leur qualité. Non seulement cette méthode

1. Dans le domaine de l'optimisation, l'utilisation du terme paramètre fait référence à un terme fixé et il serait plus juste ici de dire variable. Cependant, en hydrologie, les variables sont généralement liées à des processus hydrologiques et sont générées par la simulation plutôt qu'optimisées. Afin de ne pas créer de confusion, les termes hydrologiques seront conservés tout au long de ce document.

était-elle fastidieuse, mais seuls certains ensembles de paramètres pouvaient être étudiés sur la base de quelques rares indicateurs dû à la lourdeur de la tâche. Les résultats obtenus étaient donc loin de pouvoir être considérés comme optimaux. Depuis maintenant plus de deux décennies, le problème du calage de ces modèles a été porté à un autre niveau avec l'utilisation d'algorithmes d'optimisation et avec l'avancement des technologies permettant une augmentation considérable de la puissance de calcul. Il est maintenant impensable de paramétriser un modèle hydrologique sans faire appel aux méthodes d'optimisation. Le parallélisme, l'optimisation multi-objectif, l'optimisation séquentielle, pour ne nommer que ceux-là, se sont révélés des outils précieux et ont permis d'amener ces simulations hydrologiques à un autre niveau.

Malgré tout, plusieurs difficultés mentionnées dans Duan *et al.* (1992) font qu'au bout de plus de trois siècles d'utilisation des modèles hydrologiques, le problème n'est pas réglé et le jeu de paramètres optimal n'a toujours pas été trouvé. Notamment, la forme de la fonction, non-convexe, non-lisse et couverte de vallées et de minima locaux, constitue un défi de taille pour les algorithmes d'optimisation. Certains ont d'ailleurs été comparés dans plusieurs études, notamment celles de Goswami et O'Connor (2007), Gupta *et al.* (1999), Awad et Von Poser (2010) et Yang et Liu (2009), afin d'en identifier un qui sortirait du lot. Sans pouvoir démontrer scientifiquement qu'une méthode se démarque hors de tout doute, la majorité de ces comparaisons favorisent certaines grandes familles d'algorithmes, comme les algorithmes génétiques et les algorithmes sans dérivées. Ces derniers sont d'ailleurs conçus expressément pour le type de problèmes dont il est question ici, soit une simulation informatique impliquant une fonction potentiellement bruitée, non-lisse et non différentiable. Ces simulations deviennent rapidement coûteuses en terme de temps de calcul et l'objectif des algorithmes sans dérivées est de parvenir à la meilleure solution possible dans un contexte d'appels à la fonction limités.

1.1.1 Fonction objectif

Que ce soit en mode mono-objectif ou multi-objectif, les fonctions privilégiées par les utilisateurs pour calibrer un modèle hydrologique sont généralement des mesures d'erreur élaborées autour de la valeur du débit. Un article publié en 1998 (Gupta *et al.* (1998)), et qui reste d'actualité, résume les fonctions les plus couramment utilisées dans la littérature, comme le critère de Nash-Sutcliffe, le biais, l'écart sur les volumes de crues, pour ne nommer que ceux-là. Des onze équations présentées, la totalité implique le débit d'une façon ou d'une autre. En misant ainsi sur une seule donnée, beaucoup d'informations, qui pourraient être

mises à profit, sont ignorées. Ainsi, peu d'attention est mise sur les variables intermédiaires des processus hydrologiques, car peu d'observations sont disponibles. N'étant ni contraints, ni impliqués directement dans la fonction objectif, la modélisation des variables intermédiaires est laissée à elle-même résultant en une compensation globale des processus hydrologiques souvent peu conforme aux lois de la physique. Cela s'observe par une même valeur de la fonction objectif pour plusieurs ensembles de paramètres diamétralement opposés, ce que Beven (2006) appelle équifinalité. L'objectif d'une bonne paramétrisation étant la capacité à représenter adéquatement le régime hydrologique résultant de données n'ayant pas servi à l'optimisation, un ensemble de paramètres ne respectant pas la mécanique interne du cycle de l'eau, mais parvenant à une représentation acceptable des débits ne devrait pas être retenu.

1.2 Objectifs de recherche

L'intérêt grandissant des chercheurs pour les changements climatiques ne simplifie pas la chose : les modèles hydrologiques étant paramétrés avec les données historiques disponibles, ils ne sont pas nécessairement entraînés pour répondre aux conditions météorologiques potentiellement différentes d'un contexte de réchauffement planétaire. Cependant, la demande est bien présente, comme l'illustre la mise en place du projet $(CQ)^2$ par Hydro-Québec, en collaboration avec le Consortium Ouranos sur la climatologie régionale et l'adaptation aux changements climatiques, le Centre d'expertise hydrique du Québec (CEHQ) et l'entreprise Rio Tinto Alcan. L'objectif poursuivi par ce regroupement de chercheurs est de parvenir à estimer les impacts des changements climatiques sur le régime hydrologique québécois par le biais d'un modèle hydrologique et de multiples scénarios de changements climatiques. Le développement d'une méthode de calage plus robuste, orientée vers la reproduction des processus physiques, notamment de l'évapotranspiration, était donc d'une importance cruciale et c'est dans ce cadre que s'inscrit cette recherche.

1.3 Calage par contraintes

La méthode développée dans ce mémoire vise à répondre à cette demande de projections hydrologiques robustes en contexte de changements climatiques, mais aussi à raffiner les prévisions produites en diminuant leur variabilité et en améliorant la modélisation des processus hydrologiques afin de les rendre plus près de la physique. L'approche présentée repose sur l'hypothèse selon laquelle les données hydrométéorologiques représentent le comportement hydrologique du bassin versant, hypothèse qui sera détaillée dans la Section 4.3.3. Dans cette

optique, nous avançons une technique de calage sous contraintes qui viendra forcer le modèle HSAMI à imiter le comportement hydrologique du bassin versant en modifiant différents processus internes contrôlant les variables hydrologiques. Notre hypothèse de recherche principale est qu'une solution avec une valeur optimale de la fonction objectif qui ne respecte pas les contraintes basées sur la physique des processus est moins intéressante qu'une solution moins bonne en terme d'objectif mais qui modélise les processus hydrologiques en regard des contraintes physiques. Notre méthode devrait donc améliorer la robustesse du modèle ainsi que sa stabilité, tout en étant transposable en contexte de changements climatiques.

1.3.1 Le processus de l'évapotranspiration

Dans le cadre de ce projet, nous avons concentré nos efforts sur le processus de l'évapotranspiration pour plusieurs raisons, notamment l'importance de ce processus dans le calcul des prévisions du modèle en contexte de changements climatiques (voir Section 3.1). Cependant, comme il sera détaillé dans la Section 6.3 présentant les travaux futurs, plusieurs autres processus hydrologiques pourront être contraints selon la même approche afin d'améliorer encore la confiance accordée aux projections du modèle.

Selon Anctil *et al.* (2005), l'évapotranspiration se définit comme étant « tous les processus par lesquels l'eau gagne l'atmosphère, [...] à la fois l'évaporation de l'eau de surface et la transpiration des plantes ». Toujours selon ces auteurs, l'évapotranspiration dépend à la fois du climat, de la couverture végétale et des caractéristiques du sol. Ce processus ne peut cependant pas être mesuré directement (Beven (2001)), et c'est ce qui a limité plusieurs chercheurs qui tentaient de mettre en place une méthode comme celle présentée ici. Notamment, à la page 293 de son article, Beven (1996) mentionne qu'une validation des variables internes serait un test beaucoup plus puissant qu'une simple validation sur un ensemble de données différentes, pratique courante en hydrologie. Cependant, selon lui, il n'est malheureusement pas possible de le faire dû à une absence de données comparables. Pour un modèle simulant non seulement un débit à l'exutoire, mais aussi d'autres types de flux tel que l'évapotranspiration, Gupta *et al.* (1998) proposent de se servir de données observées, si elles sont disponibles et comparables, afin de calibrer ledit modèle. Cooper *et al.* (2007) proposent de resserrer l'espace de recherche à l'aide de contraintes afin d'orienter l'optimisation. Tous ces auteurs n'indiquent cependant pas quels types de données pourraient être utilisées ni comment procéder et la plupart précisent qu'il n'est pas possible d'y arriver avec les données disponibles. C'est donc le développement ainsi que l'application de ces idées qui sont mis de l'avant dans cette recherche.

1.3.2 Description sommaire de la méthode

Deux stratégies de calage sont mises en place et comparées à la méthode actuellement utilisée à l'aide de plusieurs critères mesurant la robustesse de la paramétrisation. La première se sert de l'Atlas du Canada pour recentré le bilan de l'évapotranspiration réelle, alors que la seconde met à profit le cycle de l'évapotranspiration réelle tel que produit par un modèle climatique. Ces deux stratégies de calage, ainsi que la stratégie employée actuellement chez Hydro-Québec sont détaillées dans la Section 4.3.

La recherche s'est déroulée naturellement en plusieurs étapes, puisque les analyses ont nécessairement dues être réalisées en ordre pour être conséquentes. En premier lieu, les contraintes devant être ajustées ont fait l'objet d'analyses de sensibilité afin d'être adaptées au problème en question. Lorsque ces analyses ont été complétées, plusieurs optimisations ont pu être comparées afin de fixer les options de l'optimiseur. Finalement, des analyses hydrologiques diverses ont été réalisées en collaboration avec une hydrologue afin de déceler certaines anomalies du processus de modélisation et afin d'évaluer la performance des méthodes. Ces analyses ont mené à revoir les stratégies de calage initiales, ce qui nous a forcé à reprendre le processus du début. Lorsque la formulation de la méthode a été considérée comme satisfaisante, plusieurs analyses liées à la robustesse des prévisions et à la stabilité des paramètres ont été effectuées, pour finalement permettre de commenter la performance des différentes stratégies et de la méthode de calage par contraintes proposée dans le cadre de ce projet.

L'ensemble des optimisations effectuées au cours de cette recherche ont été exécutées par CASIR, le centre de calcul de l'IREQ. Ce superordinateur de 1 000 cœurs a permis non seulement de lancer plusieurs optimisations simultanément, mais aussi de paralléliser chacune des optimisations de façon à accélérer le temps de calcul. Au total, ce sont des années de calcul continu qui ont été réalisées dans le cadre de ce projet.

1.3.3 L'importance des outils d'optimisation

Comme il a été mentionné plus tôt, les algorithmes d'optimisation sont devenus indispensables pour l'automatisation du calage des modèles hydrologiques. L'algorithme de recherche directe MADS, décrit à la Section 2.1.2, sera donc utilisé afin d'optimiser la paramétrisation

du modèle hydrologique HSAMI sous chacune des trois stratégies de calage. Un tel algorithme, prévu pour obtenir de bonnes solutions en peu d'évaluations de la fonction objectif, se révélera essentiel dans la réalisation de ce projet nécessitant des milliers d'optimisations indépendantes.

Le point central de la méthode développée dans ce mémoire étant l'ajout de contraintes non-lisses, une autre avancée du domaine de l'optimisation s'avérera incontournable. Récemment ajouté en complément à l'implémentation de MADS (NOMAD), le module d'analyse de sensibilité, décrit à la Section 2.2.3, permet à l'utilisateur de visualiser la sensibilité de la fonction objectif à la modification du membre de droite d'une contrainte. Cet outil viendra affiner les stratégies de calage présentées en assurant une formulation idéale des contraintes compte tenu de la sensibilité de l'objectif. À chaque modification du problème original, les contraintes pourront donc être retraitées afin de mesurer la réaction de l'objectif à un relâchement ou à un resserrement de chacune des contraintes, incluant les bornes des paramètres. L'information fournie par un tel outil est donc inestimable pour une application comme celle-ci et ouvre de nouveaux horizons quant à l'analyse de l'espace paramétrique.

1.4 Plan du mémoire

Afin de placer les lecteurs de toutes disciplines sur un pied d'égalité, les concepts de base nécessaires à la compréhension de ce projet seront exposés. Ainsi, le Chapitre 2 présentera les notions d'optimisation mises à profit dans cette recherche, notamment l'algorithme MADS à la Section 2.1.2 et le module d'analyse de sensibilité à la Section 2.2.3. Le Chapitre 3 quant à lui se consacrera au domaine de l'hydrologie en présentant les impacts possibles des changements climatiques (Section 3.1), les modèles climatiques (Section 3.2) et les modèles hydrologiques (Section 3.3). Suivra, au Chapitre 4, la description détaillée du problème d'optimisation, incluant les stratégies de calage proposées à la Section 4.3. Les résultats dans leur intégralité seront présentés au Chapitre 5 et des résultats complémentaires réalisés en collaboration avec un groupe d'hydrologues sont présentés en Annexe G afin de permettre une généralisation des conclusions. Finalement, au Chapitre 6, une synthèse du projet sera produite et les contributions de la méthode, ses limitations ainsi que certaines améliorations futures seront exposées.

CHAPITRE 2

REVUE DE LITTÉRATURE

Cette section se veut une présentation télescopique des principaux éléments de recherche opérationnelle ayant trouvé usage dans cette recherche. Ainsi, l'horizon le plus large sera consacré à l'historique des méthodes de recherche directe ayant conduit à l'algorithme de recherche directe sur treillis adaptatif (*Mesh Adaptative Direct Search*) (MADS). Celui-ci sera ensuite examiné plus en profondeur et ses principales caractéristiques seront présentées. En se rapprochant encore, l'implémentation de MADS, NOMAD, sera décrite ainsi que ses outils complémentaires de gestion de contraintes.

2.1 Évolution des méthodes de recherche directe

Depuis l'apparition du calcul différentiel, au XVIIe siècle, la recherche des optimums d'une fonction a envahi les pensées de nombreux mathématiciens. Cependant, nous avons dû attendre jusqu'en 1947 pour que Dantzig (1951) introduise le terme programmation linéaire et que cette branche des mathématiques soit traitée comme un domaine scientifique à part entière. À partir de ce moment, les méthodes de résolution du problème type d'optimisation [1],

$$\min_{x \in \Omega} f(x) \tag{2.1}$$

où $\Omega = \{x \in X : c_j(x) \leq 0,\ j \in J \subset \mathbb{R}^n\}$, $f,\ c_j : X \to \mathbb{R} \cup \{\infty\}$ pour tout $j \in J = \{1, 2, ..., m\}$, et où X est un sous-ensemble de $\mathbb{R}^n$,

se sont diversifiées afin de mieux répondre à certains besoins (résolution en nombres entiers, résolution de problèmes de grande dimension, etc.).

La recherche directe a pour particularité de ne nécessiter aucun calcul exact ni aucune estimation du gradient de la fonction à optimiser. Les algorithmes de cette famille de méthodes

1. Notons que la maximisation d'une fonction est équivalente à la minimisation de son opposée ($\max\ f(x) \equiv -\min\ -f(x)$). Ainsi, cette formulation du problème d'optimisation n'est pas restrictive.

d'optimisation numérique sont donc une option des plus judicieuses pour l'optimisation de fonctions non-linéaires très bruitées, de fonctions non lisses ou non différentiables, mais aussi dans le cas d'une simulation ou d'une boîte noire où, par souci de confidentialité ou par impossibilité d'accéder à l'implémentation du problème, seules les données fournies en entrée ainsi que les sorties sont accessibles à l'optimiseur.

2.1.1 Recherche par coordonnées et recherche par motifs généralisée

C'est en 1950 que les algorithmes de recherche directe sur treillis voient le jour avec la recherche par coordonnées (CS) (brève description en préface de Davidon (1991)). Le principe général consiste à se déplacer sur une grille virtuelle dans l'espace des variables appelée treillis (en anglais, *mesh*) en évaluant la valeur de la fonction objectif dans certaines directions et en se déplaçant dans le sens de l'amélioration. Le raffinement de la grille variera alors tout au long de l'optimisation selon des règles bien précises liées aux résultats obtenus. Le concept de recherche sur treillis est donc introduit et n directions de recherche de base et leur n inverse, soit $2n$ directions, sont exploitées. Par la suite, en 1996, Torczon crée une généralisation de cet algorithme, la recherche par motifs généralisée (*Generalized Pattern Search*) (GPS), qui donnera accès à un plus grand nombre de directions de recherche.

2.1.2 L'algorithme MADS

MADS, algorithme de recherche directe sur treillis adaptatif (*Mesh Adaptative Direct Search*), s'inspire de l'algorithme GPS (recherche par motifs généralisée ou *Generalized Pattern Search*), que l'on doit à Booker *et al.* (1999), qui eux-mêmes se sont inspirés des travaux de Torczon (1997). La première instantiation, MADS triangulaire inférieur (*Lower Triangular MADS*), offre un nombre infini de directions de recherche, mais qui ne sont pas générées par groupes orthogonaux, ce qui entraîne une perte de certaines parties de l'espace de recherche. De plus, LT-MADS est probabiliste, donc les résultats ne peuvent pas être répétés d'une optimisation à l'autre. Afin de corriger ces lacunes, une deuxième version, OrthoMADS (Abramson *et al.* (2009)) a été lancée, offrant tous les avantages de la première instantiation, sans les inconvénients.

OrthoMADS génère un ensemble de directions dense dans l'espace, comme LT-MADS, mais sans l'aspect aléatoire. En effet, les directions de Ortho-MADS sont générées à l'aide de la suite quasi-aléatoire de Halton, ce qui permet de conserver l'ensemble de directions

d'une optimisation à l'autre. Toutefois, il est aussi possible de modifier cet ensemble en fournissant un paramètre, la graine aléatoire (ou *seed*), à l'algorithme. Un autre avantage est lié aux directions de recherche qui sont orthogonales entre elles, ce qui vient maximiser la couverture de l'espace à chaque itération. Les preuves liées à la convergence des algorithmes GPS et MADS ont été produites par Audet et Dennis, Jr. (2003), Audet (2004), Audet et Dennis, Jr. (2006), Audet *et al.* (2008b) et Vicente et Custódio (2012), en se basant sur le calcul non lisse de Clarke (1983). L'algorithme MADS, étant un algorithme de recherche directe, ne nécessite aucune évaluation de dérivée et n'impose aucune condition de différentiabilité sur la fonction à optimiser. Il était donc tout indiqué pour le problème de type boîte noire impliquant une simulation complexe et bruitée auquel nous avons affaire dans ce projet.

Description de MADS

De façon générale, l'algorithme MADS, à partir d'un point de départ x_0, se déplace sur une grille virtuelle appelée treillis, en avançant d'un point réalisable à l'autre dans le sens de l'amélioration. La structure de l'algorithme, représentée par le Schéma 2.1, se décline en trois étapes : l'initialisation, la recherche et la sonde, et finalement, les mises à jour.

Initialisation En premier lieu, les paramètres tels que le point de départ (x_0), le paramètre de taille de sonde (Δ_k^p), ainsi que le paramètre de taille du treillis (Δ_k^m) sont initialisés. Le paramètre de taille de sonde Δ_k^p vient délimiter une zone de sonde autour du point courant x_k. Celui-ci devient alors le centre du cadre de recherche autour duquel un ensemble de directions sera généré. Le paramètre de taille du treillis, ou Δ_k^m, quant à lui, indique le niveau de raffinement du treillis et sera toujours inférieur ou égal au paramètre de taille de sonde ($\Delta_k^m \leq \Delta_k^p$). L'implémentation de MADS, NOMAD, est conçue de telle façon que des valeurs par défaut pour les paramètres associés au treillis seront fournies, mais l'usager se doit de procurer à l'optimiseur un point de départ valide en regard des contraintes non relaxables (voir la section 2.2.1 pour la définition des types de contraintes). Cette étape d'initialisation n'a lieu que lors du lancement de l'optimisation et ne sera pas reprise à chaque itération de l'algorithme, contrairement aux étapes suivantes.

Recherche et sonde Ensuite, l'étape de recherche, s'il y a lieu, ainsi que l'étape de sonde sont réalisées. L'étape de recherche est une étape optionnelle. À chaque itération, elle permet aux connaisseurs d'un modèle d'exécuter une recherche dans une zone de l'espace paramétrique qu'ils jugent prometteuse, sans nuire aux preuves de convergence de l'algorithme, à la condition qu'un nombre fini d'évaluations soient réalisées et que celles-ci se situent sur le

treillis. Cette étape est aussi l'occasion d'utiliser une fonction substitut du modèle (Booker *et al.* (1999), Conn et Le Digabel (2011), Gramacy et Le Digabel (2011)) ou d'intégrer une métaheuristique (voir la Section 2.1.2). L'étape de sonde quant à elle est une recherche locale de l'espace. À partir du centre de sonde courant x_k et du paramètre de taille de sonde Δ_k^p, le voisinage de sonde sera défini par un ensemble de directions D_k formant une base positive. De ce voisinage, des points d'essais, nécessairement situés sur le treillis, seront extraits puis soumis à la boîte noire pour évaluation. Dans un premier temps, le respect des contraintes gérées par la barrière extrême (définie à la Section 2.2.2) sera vérifié. Dans le cas où un point d'essai ne satisferait pas ces conditions, il serait automatiquement rejeté, sans considérer la fonction objectif qui sera fixée à $+\infty$. Si le point $y \in X$, autrement dit si le point respecte les contraintes non relaxables, alors un appel complet à la boîte noire sera effectué et la fonction objectif ainsi que les autres contraintes seront évaluées. Cette façon de faire est conçue afin d'épargner à l'utilisateur le temps de calcul nécessaire à une évaluation de la boîte noire, parfois très coûteuse, pour un point candidat qui ne sera, quoi qu'il arrive, pas retenu. Il peut arriver toutefois que la boîte noire sera définie de telle façon qu'un appel complet sera nécessairement effectué. Dans ce cas, la fonction objectif sera tout de même fixée à $+\infty$, mais le coût de l'appel à la fonction, en termes de temps de calcul, ne sera pas réduit. Notons aussi qu'une base de données appelée la *cache* tient l'historique des points évalués. Ainsi, avant toutes choses, un point d'essai sera recherché dans la *cache* et ne sera pas resoumis à la boîte noire s'il devait être trouvé.

Mises à jour Selon le dénouement de l'étape de sonde, les paramètres seront mis à jour en suivant différentes règles. Deux cas de figure sont possibles et chacun est lié à son propre protocole d'adaptation des paramètres. Ainsi, dans le premier cas, un point candidat de l'étape de recherche ou de sonde a une valeur de l'objectif inférieure à la solution courante x_k. Ce point deviendra alors le nouveau centre de sonde et la taille du treillis ainsi que le paramètre de sonde seront augmentés ou conservés. L'algorithme considérera qu'il se situe dans une pente descendante et, sous la stratégie optimiste, prolongera alors la longueur de ses pas afin d'accélerer la convergence. Dans le second cas, jugé un échec, aucun point d'essai n'offrira une meilleure performance que la solution courante. En conséquence, l'algorithme évaluera le voisinage du point courant comme se situant dans une vallée et les paramètres de sonde et de taille du treillis seront réduits afin d'en trouver l'optimum, qui pourrait être global ou local. Lors de l'optimisation d'un problème contraint, certaines subtilités viennent se greffer à ces règles de mises à jour des paramètres. Celles-ci seront détaillées à la Section 2.2.

Étape 1: Initialisation

1. Choix d'un point de départ x_0 tel que $x \in X$
2. Définition d'un ensemble de directions D formant une base positive
3. Initialisation des tailles de sonde et de treillis : $0 < \Delta_0^m < \Delta_0^p$
4. Initialisation du compteur $k \leftarrow 0$

Tant qu'un critère d'arrêt n'est pas satisfait, répéter les étapes 2 à 4

Étape 2 : Recherche (optionnelle)

1. Générer un ensemble fini de points y sur le treillis M_k
2. Pour chaque point y, évaluer la fonction ainsi que les contraintes
3. Si un point y est meilleur que x_k, passer directement à l'étape 4
4. Sinon, passer à l'étape 3

Étape 3 : Sonde

1. Générer un ensemble de directions D_k
2. Construire l'ensemble des points de sonde $P_k = \{x_k + \Delta_{k+1}^m d : d \in D_k\} \subset M_k$
3. Évaluer la fonction ainsi que les contraintes pour les points de P_k

Étape 4 : Mises à jour

1. Si un point y est meilleur que x_k :
    a. $x_k \leftarrow y$
    b. Mettre à jour $\Delta_{k+1}^m \geq \Delta_k^m$ et $\Delta_{k+1}^p \geq \Delta_{k+1}^m$
2. Sinon :
    a. $x_{k+1} = x_k$
    b. Mettre à jour $\Delta_{k+1}^m < \Delta_k^m$ et $\Delta_{k+1}^p \geq \Delta_{k+1}^m$
3. $k \leftarrow k + 1$

Figure 2.1 Formulation de l'algorithme MADS.

Fonctionnalité de MADS

À l'algorithme MADS de base se sont greffées plusieurs fonctionnalités au fil du temps, notamment la mise à profit de la parallélisation des tâches, l'optimisation bi-objectif avec BiMADS (Audet *et al.* (2008d), Audet *et al.* (2010b)), l'option VNS_SEARCH (Audet *et al.* (2008a)) qui effectue une recherche à l'aide de la métaheuristique de recherche à voisinage variable (*Variable Neighborhood Search*) et plusieurs autres qui ne seront pas détaillées ici.

MADS avec décomposition parallèle de l'espace (*Parallel Space Decomposition with MADS*) (PSD-MADS) (Audet *et al.* (2008c)) La résolution de très larges problèmes est souvent limitée par la puissance informatique disponible, notamment par le temps de calcul nécessaire. Afin de contourner cette difficulté, et lorsque l'installation informatique le permet, l'utilisation de multiples processeurs simultanément entraîne une division de la charge de travail et, donc, une accélération globale de la résolution. On fait référence à cette

façon de faire comme le parallélisme ou la parallélisation des tâches.

La technique utilisée par MADS pour la décomposition de l'espace en parallèle figure dans son intégralité dans Audet *et al.* (2008c) et sera résumée dans cette section. L'idée générale consiste à faire travailler chaque processeur sur un sous-ensemble de variables. Les variables de chaque ensemble seront choisies aléatoirement et celles ne faisant pas partie de l'ensemble seront temporairement fixées à leur dernière valeur, soit celle de la solution courante (une variante est aussi proposée par Yahia (2012), où les variables sont choisies selon leur impact statistique sur la fonction objectif). Ainsi, à chaque itération, un processeur maître divisera le problème principal en sous-problèmes qui seront assignés à des processeurs esclaves. Dans le cas où le nombre de sous-problèmes serait supérieur aux processeurs disponibles, les tâches restantes seront distribués aux premiers processeurs libérés. Ceux-ci résoudront le problème qui leur est dévolu de façon totalement indépendante, puis les résultats seront retournés au maître qui les comparera entre eux en temps réel. Dès qu'une solution inférieure à la solution courante sera reçue, celle-ci deviendra le nouveau itéré courant x_k et les paramètres de taille de sonde et de treillis seront actualisés. Notons que les règles concernant l'optimisation contrainte s'appliquent toujours. Une solution présentant une violation de contrainte moindre sera donc aussi considérée comme une amélioration. Les tâches en cours seront tout de même menées à bien, mais seront confrontées à la nouvelle solution courante. L'ordre selon lequel les tâches sont retournées au maître peut donc influer sur la solution finale, comme il sera discuté dans la section 4.2.1. Il est intéressant de noter que les analyses de convergence menées pour l'algorithme MADS sont toujours valides dans un contexte d'utilisation parallèle.

BiMADS L'algorithme MADS a aussi été adapté afin de pouvoir résoudre les problèmes d'optimisation comportant plusieurs objectifs conflictuels. Lorsqu'un tel cas se présente, le déroulement de l'algorithme est exactement le même ; c'est le problème qui sera modifié et non l'inverse. Ainsi, une méthode a été développée en 2008 (Audet *et al.* (2008d)) permettant de reformuler un problème multiobjectif en une série de reformulations mono-objectif, qui sont résolues par l'algorithme MADS tel qu'on le connaît. Cette fonctionnalité n'ayant pas été mise à profit dans le cadre de ce projet, la théorie soutenant la validité de la reformulation, ainsi que les preuves de convergence, ne seront pas évoquées.

VNS L'utilisation de la VNS, quant à elle, a joué un rôle très important dans la résolution de notre problème. Une attention particulière lui sera donc accordée. C'est donc en 2006 que la métaheuristique de Mladenović et Hansen (1997), la recherche à voisinage variable (*Variable Neighborhood Search*), fut couplée à l'algorithme MADS afin de permettre une

meilleure exploration de l'espace. L'intérêt de l'utilisation de cette méthode réside dans sa capacité à s'échapper de la zone d'attraction d'un optimum local et dans sa faculté de passer d'une section à l'autre d'un domaine fragmenté. La structure de MADS est telle que l'ajout de cette méthode de recherche allait de soi, et ce sans altérer les preuves de convergence de l'algorithme. En effet, le déroulement de MADS se découpe en deux étapes, soit la recherche et la sonde, tel que mentionné plus haut. L'étape de recherche très flexible, se prête à merveille à l'utilisation de la VNS. En effet, les seules conditions imposées sur les points explorés au cours de cette étape sont qu'ils se situent sur le treillis et qu'un nombre fini de points soit généré.

La recherche à voisinage variable (*Variable Neighborhood Search*) (VNS) à proprement parler est définie par deux concepts : soit une méthode de descente et la définition d'un voisinage (voir Hansen et Mladenović (2001) pour les détails). Globalement, la recherche s'effectue en explorant le voisinage d'un point et en se déplaçant en direction de l'amélioration, selon les instructions de la méthode de descente. Le point servant de centre pour la recherche est obtenu par perturbation plus ou moins importante du meilleur point courant ; perturbation qui peut être définie différemment selon le problème, ou, dans le cas qui nous intéresse, qui peut être adaptée au type d'optimiseur. Ainsi, afin d'exploiter au mieux la séquence d'optimisation de MADS, la perturbation a été définie de façon à ce que, en cas d'échec lors de l'étape de sonde, l'exploration de l'espace par la VNS et par MADS se complètent. MADS viendra raffiner la taille de son treillis, alors que le voisinage du point exploré par la VNS s'élargira.

L'utilisation de cette fonctionnalité implique de fournir à l'optimiseur une valeur complémentaire correspondant au pourcentage du nombre total d'évaluations de MADS qui seront utilisées par la VNS. Par exemple, en indiquant un pourcentage de 90, sur le nombre total d'appels à la boîte noire, supposons 20 000, 90 % de celles-ci auront été exécutées par la VNS, donc 18 000, et le reste sera accordé à la recherche de MADS, soit 2 000. La VNS étant lancée à chaque étape de recherche précédant chaque étape de sonde, la division des évaluations se fait de façon assez exacte, à quelques évaluations près. Le choix du degré d'utilisation de la VNS pour un problème donné ne peut être déduit par de quelconques formules et doit nécessairement être testé. Ainsi, la section 5.1.2 détaillera la méthode utilisée dans le cadre de cette recherche pour arriver à une valeur optimisant l'utilisation de cette métaheuristique.

2.1.3 Le logiciel NOMAD

Pour mener à bien l'optimisation de notre problème, la dernière version de l'optimiseur NOMAD (Nonsmooth Optimization by Mesh Adaptative Direct Search) (Le Digabel (2011)), une implémentation en C++ de l'algorithme de recherche directe sur treillis MADS, fut utilisée. Les raisons sous-tendant ce choix sont nombreuses et variées. En premier lieu, l'algorithme derrière cet optimiseur a été conçu spécialement pour des problèmes comme celui dont il est question ici, comme il fut mentionné dans la section 2.1.2. De plus, NOMAD était déjà employé par Hydro-Québec pour la calibration du modèle hydrologique visé par cette étude et avait des performances comparables aux autres algorithmes disponibles. La raison principale de ce choix, par contre, est basée sur l'utilisation du module complémentaire d'analyse de sensibilité, qui sera décrit plus loin (à la Section 2.2.3). Puisque l'intérêt de ce travail repose sur l'ajout de contraintes au modèle hydrologique, l'étude de la sensibilité de la fonction objectif à chacune desdites contraintes est primordiale. Malgré le fait que le module soit indépendant de NOMAD, il était plus facile et plus naturel de l'utiliser avec l'optimiseur pour lequel il avait été conçu en premier lieu.

2.2 La gestion des contraintes

Puisque l'ajout et la gestion des contraintes sont au coeur de ce projet, une attention particulière sera portée à ce sujet. Afin de bien saisir l'utilisation des outils d'analyses de sensibilité offerts par NOMAD, il est important de comprendre la façon dont les contraintes sont gérées par MADS, qui est dépendante de la classe à laquelle elles appartiennent.

2.2.1 Les types de contraintes

Un problème d'optimisation peut compter jusqu'à trois types de contraintes : les contraintes non relaxables, les contraintes relaxables ainsi que les contraintes cachées. La première catégorie contient toutes les contraintes ne pouvant être violées dans aucune situation, sans quoi la simulation ne pourra pas être exécutée, ou n'aura aucun sens. Par exemple, une mesure de longueur doit nécessairement être positive. La seconde catégorie, généralement de la forme $c(x) \leq 0$, contient une série de contraintes pouvant être relâchées ou resserrées, et qui peuvent être violées en cours d'optimisation, pour autant qu'elles soient respectées à la solution finale. Finalement, les contraintes cachées ne sont pas connues au moment de l'optimisation. Elles sont considérées responsables lorsqu'un point respectant toutes les autres contraintes est évalué et que la simulation échoue à retourner une valeur.

2.2.2 Les méthodes de gestion des contraintes

La barrière extrême

Ces contraintes étant très différentes, plusieurs techniques de gestion de contraintes ont été développées. La première est celle de la barrière extrême (*Extreme Barrier*) (EB), adaptée pour MADS par Audet *et al.* (2010a). Cette méthode consiste à fixer l'objectif à l'infini lorsque les points sont irréalisables, ce qui revient à traiter un problème non contraint. La barrière extrême est utilisée pour les contraintes non relaxables, car, par définition, elles doivent être respectées à chaque évaluation, sans quoi la solution n'a aucun sens, et donc aucune valeur opérationnelle. L'utilisation de cette technique implique que le point de départ doit satisfaire ces contraintes, sans quoi l'optimisation ne pourra être réalisée. Mathématiquement, cette stratégie se définit comme suit :

$$f_\Omega(x) = \begin{cases} f(x) & \text{si } x \in \Omega, \\ \infty & \text{sinon.} \end{cases} \tag{2.2}$$

Le filtre

En 2002, la méthode du filtre, plus permissive, a été introduite par Fletcher et Leyffer (2002) pour être ajoutée, après quelques légères modifications, aux fonctionnalités de MADS en 2006 par Audet et Dennis, Jr. (2004). Ce type de gestion des contraintes quantifiables dites relaxables permet l'exploration de points non réalisables en cours d'optimisation afin de raffiner la recherche. Cette méthode repose sur un sous-problème bi-objectif, minimisant à la fois la fonction objectif du problème principal ainsi qu'une fonction de mesure de violation, $h(x)$, agrégeant la violation de toutes les contraintes traitées avec cette méthode. L'emphase est mis sur la fonction de violation de contraintes afin de pouvoir éventuellement converger vers des solutions réalisables. Un point sera accepté comme faisant partie du filtre à la condition d'améliorer l'un ou l'autre des deux objectifs. En fait, le filtre principal ne contient que des points irréalisables afin de ne pas bloquer la recherche, tandis que le meilleur point réalisable, quant à lui, formera un filtre d'un seul point et sera remplacé si une meilleure solution venait à être trouvée. Chaque point d'essai sera donc comparé au filtre courant et sera rejeté ou viendra s'ajouter au filtre, en remplissant une ouverture ou en remplaçant un point faisant déjà partie du filtre. La fonction de pénalité $h(x)$ se calcule de la façon suivante :

$$h(x) = \begin{cases} 0 & \text{si } x \text{ est un point réalisable,} \\ \sum_{j=1}^{m}(\max(0, c_j(x)))^2 & \text{sinon,} \end{cases} \tag{2.3}$$

Le principal avantage du filtre sur les autres méthodes du même type est qu'il ne dépend d'aucun paramètre évaluant la violation de contrainte de façon arbitraire. Les résultats obtenus ne sont donc pas liés à un choix de l'utilisateur.

La barrière progressive

Ensuite, la méthode de la barrière progressive (*Progressive Barrier*) (PB) proposée par Audet et Dennis, Jr. (2009) est venue remplacer le filtre. Cet algorithme utilise le même type de fonction mesurant la violation des contrainte, la différence étant qu'un seuil h^{max}, monotone décroissant, vient contraindre cette fonction de façon à ce que les points dont la violation est trop importante soient rejetés. Ce seuil sera resserré de façon progressive afin de converger vers une solution réalisable, tout en permettant une meilleure exploration de l'espace que la méthode précédente. Le seuil initial, h_0^{max}, peut être imposé par l'utilisateur ou la valeur par défaut de NOMAD (∞) peut être utilisée. Le calcul de la violation de contraintes se fait comme suit :

$$h(x) = \begin{cases} \sum_{j \in J} (\max(c_j(x), 0))^2 & \text{si } x \text{ est un point respectant} \\ & \text{les contraintes de type EB,} \\ \infty & \text{sinon .} \end{cases} \tag{2.4}$$

Pour l'application de cette méthode, plusieurs ensembles seront définis afin de classer les points testés. L'ensemble faisable courant, F_k, contiendra la meilleure solution réalisable, tandis qu'un deuxième ensemble, U_k, contiendra les points irréalisables non dominés au sens de Pareto, ce qui signifie que l'objectif ne peut être amélioré sans que la violation des contraintes ne soit dégradée ou inversement. De cet ensemble découlera le sous-ensemble irréalisable courant, I_k, qui sera constitué des points faisant partie de l'ensemble U_k et respectant le seuil h^{max}. À partir du moment où un point est inscrit dans cet ensemble, le sous-ensemble I_k ne sera plus jamais vide grâce aux règles imposées sur h^{max}.

Les solutions considérées à l'itération k correspondent aux points contenus dans les ensembles F_k et I_k ayant chacun la meilleure valeur de fonction-objectif de leur ensemble. Deux ensembles distincts seront aussi considérés durant l'étape de sonde afin d'explorer à la fois l'espace réalisable, ou respectant la totalité des contraintes, ainsi que l'espace des points violant les contraintes relaxables et quantifiables, mais respectant celles dites non relaxables. Les efforts entre les deux centres de sonde seront répartis selon la formule suivante : si $f_k^I < f_k^F - \rho$, alors l'ensemble I_k sera considéré comme le centre primaire de sonde, alors que l'ensemble F_k

deviendra le centre secondaire et bénéficiera de moins d'attention. La valeur de ρ est fixée à 0.1 par défaut, mais peut être personnalisée par l'utilisateur en fonction du problème à optimiser.

La variation du seuil h^{max} est dépendante du résultat de la dernière itération réalisée. Trois cas de figure sont possibles : l'itération peut être un succès, un échec ou une amélioration. Dans le cas d'un succès, le point trouvé domine au sens de Pareto toutes les solutions présentes dans l'ensemble concerné, soit I_k ou F_k. La taille du treillis, Δ_k^m, sera augmentée et le seuil h^{max} sera diminué de façon à égaler la plus grande valeur de h inférieure au seuil courant. Advenant un échec, c'est-à-dire qu'aucun point trouvé n'a su améliorer la fonction objectif ou la violation de contrainte, la taille du treillis sera diminuée et le seuil restera inchangé. Finalement, une amélioration est définie comme étant un point irréalisable ayant une violation de contraintes strictement plus petite que celles de tous les points formant I_k. Dans ce cas, la taille du treillis et le seuil seront conservés.

La barrière progressive-extrême

Finalement, une méthode hybride entre la barrière extrême et la barrière progressive, la barrière progressive-extrême (*Progressive-to-Extreme Barrier*) (PEB) (Audet *et al.* (2010a)), a vu le jour afin d'offrir une alternative lorsque aucun point de départ réalisable n'est disponible, mais que la barrière extrême semble être la technique la plus appropriée. Toutes les contraintes sont initialement traitées à l'aide de la méthode de la barrière progressive. Dès qu'un point arrive à satisfaire une des contraintes, celle-ci est prise en charge par la barrière extrême. L'ensemble U_k est alors reconstruit en conséquence, puisque certains points ont maintenant une violation de contraintes fixée à l'infini. Chaque contrainte est ainsi traitée jusqu'à ce qu'une solution respectant l'ensemble des contraintes soit atteinte.

2.2.3 Analyses de sensibilité des contraintes

Lorsque des contraintes relaxables sont ajoutées à un modèle, ou lorsque la valeur du membre de droite d'une contrainte est incertaine ou modifiable, il est intéressant pour l'usager d'évaluer la pertinence de la limite choisie. Dans certains cas idéaux, où la fonction à optimiser est lisse et où les contraintes respectent certaines conditions, de telles analyses, nommées analyses de sensibilité, peuvent être réalisées à l'aide des multiplicateurs de Lagrange. Ceux-ci quantifient l'effet sur la fonction objectif d'une variation d'une unité du membre de droite de la contrainte. Malheureusement, plus souvent qu'autrement, ces mul-

tiplicateurs ne sont pas uniques ou n'existent tout simplement pas. Dans ces cas, plus près de la réalité, les analyses de sensibilité restent pertinentes et c'est pourquoi, en 2010, Audet *et al.* (2012) ont mis au point deux outils d'analyse pouvant être mis à profit par n'importe quel algorithme d'optimisation sans dérivée. Le premier produit une analyse de sensibilité plus approximatif, mais sans coût, tandis que le second nécessitera un budget d'évaluations supplémentaires, mais fournit une analyse beaucoup plus fine.

La méthode simple

À la suite d'une optimisation, NOMAD fournit un fichier de *cache* contenant l'historique des points évalués (leurs coordonnées, la valeur de l'objectif et de chacune des contraintes). L'analyse simple, comme son nom l'indique, parcourt la cache et produit un fichier contenant la valeur de la fonction objectif du problème en fonction de la valeur de la violation de la contrainte analysée. Il suffit de fournir à l'outil un fichier de *cache*, ainsi que l'index de la fonction objectif et de la contrainte à analyser. Optionnellement, il est aussi possible d'ajouter un fichier de bornes sur les variables, si l'analyse de ces contraintes particulières est pertinente. Cette analyse a l'avantage d'être gratuite en termes d'évaluations, mais elle peut être plus ou moins pertinente selon l'exploration de l'espace effectuée lors de l'optimisation soumise à l'analyse. Dans le cadre de notre recherche cependant, cet outil s'est révélé suffisant pour nos besoins.

La méthode détaillée

L'analyse détaillée, quant à elle, réalise une optimisation bi-objectif en conservant l'objectif initial et en considérant la contrainte à l'étude comme deuxième objectif. Le résultat de l'application de cet outil sera l'ensemble de points Pareto représentant le meilleur compromis entre la fonction-objectif et la violation de la contrainte. Afin d'assurer la bonne performance de l'outil d'analyse détaillée, il est nécessaire de spécifier l'index de la contrainte étudiée ainsi que les bornes d'un intervalle d'intérêt sur la violation de contrainte, ℓ_j et u_j.

L'analyse

À partir des fichiers produits par chacun des deux outils d'analyses de sensibilité, il est possible de tracer un graphique représentant l'évolution de la valeur de la fonction objectif en fonction de la valeur de la contrainte. Ce graphique sera ensuite analysé visuellement afin de

déterminer si un relâchement ou un resserrement de la contrainte est souhaitable. Les analyses effectuées par Audet *et al.* (2012) ont montré que la valeur de la pente des dits-graphiques tendait vers la valeur des multiplicateurs de Lagrange lorsque certaines qualifications des contraintes étaient satisfaites et que les fonctions étudiées étaient lisses.

CHAPITRE 3

INTRODUCTION ET APPLICATION DES CONCEPTS DE BASE EN HYDROLOGIE

Ce projet étant une application de l'optimisation au domaine de l'hydrologie, ces deux sciences se sont enchevêtrées au point de ne plus pouvoir être dissociées. Ainsi, afin que la portée de cette recherche soit accessible à tous, cette section se veut une mise à niveau destinée au lecteur désirant en connaître plus sur l'aspect hydrologique de la question. Un bref résumé des principales notions de base en changements climatiques sera présenté ici et les différents modèles, hydrologiques et climatiques, ainsi que leur utilisation dans ce projet seront survolés.

3.1 Les changements climatiques

Depuis déjà plusieurs années, les changements climatiques sont une préoccupation pour plusieurs chercheurs et entreprises. L'augmentation constante des gaz à effet de serre, et ce même dans les scénarios les plus optimistes, laisse présager une réalité météorologique future différente de ce qui est vécu actuellement. La production d'hydroélectricité étant intimement liée aux variables météorologiques, une bonne compréhension du phénomène dans le but de produire des projections fiables est nécessaire à la pérennité des entreprises de ce domaine. Cependant, comme le mentionne Salathé Jr *et al.* (2007), la transférabilité dans le futur d'un modèle hydrologique est une incertitude fondamentale. En effet, Bergström *et al.* (2002) soulèvent la question de la pertinence de projections futures réalisées à partir d'un calage sur des données historiques. Ces questionnements ont été entendus chez Hydro-Québec, d'où la conduite de cette étude.

Le Consortium Ouranos, important regroupement de chercheurs de plusieurs disciplines étudiant les changements climatiques, a produit un document en 2010 (DesJarlais *et al.* (2010)) exposant notamment les impacts appréhendés sur la production d'hydroélectricité au Québec. À l'horizon 2050, pour le nord du Québec, provenance de plus de 80 % de cette production, une hausse des températures de 4.5 à 6.5 ℃ en hiver et de 1.6 à 2.8 ℃ en été viendrait réduire les périodes de gel ainsi que la quantité et la persistance de la neige au sol. De plus, une hausse des précipitations de 16.8 à 29.4 % en hiver et de 3.0 à 12.1 % en été est projetée. Sachant que les deux principaux facteurs influençant positivement l'évapotrans-

piration sont l'augmentation des températures et de la quantité d'eau disponible dans le sol, il semble assez probable que ce processus hydrologique sera très perturbé par les bouleversements climatiques attendus. Selon Lofgren *et al.* (2011), ce processus serait appelé à jouer un rôle clé dans la modélisation d'apports réalistes pour le futur, et pourtant la différence de changements prévus pour l'évapotranspiration en contexte de changements climatiques, selon que le calcul a été réalisé par un modèle climatique ou un modèle hydrologique, est considérable. Comme le montre la figure présentée à l'Annexe A, les modèles hydrologiques simulent des valeurs étant jusqu'à trois fois plus élevées que celles des modèles climatiques (Milly et Dunne (2011)). Afin de prétendre à une projection décente des apports naturels en contexte de changements climatiques, la modélisation du processus d'évapotranspiration devait nécessairement être revu et c'est pourquoi cette variable d'état a été ciblée pour ce projet.

3.2 Modélisation climatique

C'est donc afin de permettre l'étude de différents phénomènes, principalement le réchauffement climatique, ainsi que de leurs répercussions sur l'activité humaine et les ressources naturelles que les modèles climatiques ont été élaborés. Ces modèles sont des simulations informatiques basées sur les lois physiques fondamentales visant à reproduire de façon réaliste le comportement climatique terrestre, en modélisant plusieurs phénomènes comme les échanges d'énergie entre la Terre, les océans, l'atmosphère et l'espace, les transferts radiatifs et la circulation de l'air et de l'eau dans l'atmosphère, la circulation océanique, les nuages et les échanges de carbone entre l'atmosphère et la planète. Le premier modèle atmosphérique a vu le jour en 1950, et, au décompte de 2007, c'était une vingtaine de différents modèles qui s'offraient aux chercheurs à travers le monde. Ces modèles créent un maillage de l'espace en tuiles de plusieurs kilomètres de côté, et prennent en compte la troisième dimension en superposant des boîtes afin de modéliser les interactions entre les différentes couches de l'atmosphère. Ces modèles sont d'une complexité impressionnante : selon le Réseau canadien en modélisation et diagnostics du climat régional (MDCR), pour chaque pas de temps et pour chaque point de grille, ce sont plus de 100 000 instructions qui sont exécutées. Ainsi, ce ne sont pas les connaissances qui, actuellement, limitent le développement de ces modèles, mais plutôt les ressources informatiques.

3.2.1 MRCC, modèle climatique

La version du Modèle régional du climat canadien (MRCC) telle qu'elle a été utilisée pour ce projet a été développée par (Caya et Laprise (1999)). Ce nouveau modèle était très différent de ses prédécesseurs en cela qu'il utilisait un noyau dynamique basé sur les équations du champ d'Euler, permettant une utilisation sous toutes les échelles spatiales. Ses performances sont aujourd'hui limitées aux pas de temps de quinze minutes et aux points de grille séparés d'une distance inférieure à 45 km. Contrairement aux modèles globaux, les modèles régionaux permettent de limiter la simulation du climat à une région donnée du globe, permettant des résolutions beaucoup plus fines. Par exemple, la résolution du MRCC permet la simulation de processus climatiques de petite dimension tels que le développement des nuages ou des orages, les précipitations, l'évaporation et l'humidité au sol. Ce type de modèle doit cependant être alimenté par les données de modèles de circulation générale aux frontières latérales et inférieures, technique nommée emboîtement. À partir de ces conditions initiales, le modèle régional génère ses propres structures et produit un scénario possible du climat de cette région. Ainsi, non seulement les territoires moins nantis en terme de qualité ou de quantité de données météorologiques peuvent-ils être étudiés grâce aux modèles régionaux du climat, mais la simulation du climat futur est aussi à notre portée.

Utilisation du modèle

Le MRCC version 4.2.3 a été mis à profit dans le cadre de ce projet, pour la période de simulation 1961-2001, sur un domaine centré sur le Québec (111×87 points de grille) avec résolution horizontale de 45 km. La simulation a été pilotée à ses frontières par les champs atmosphériques provenant des réanalyses globales ERA40 de l'ECMWF (European Center for Medium-Range Weather Forecasts ReAnalyses (Uppala *et al.* (2005))), publiquement disponibles sur une grille globale régulière de $2.5 \times 2.5\ ^{\circ}$ lat-lon. Le nom opérationnel de cette simulation est *aft* et sera réutilisé dans la présentation des résultats. De toutes les sorties fournies par cette simulation, seule la variable modélisant l'évapotranspiration, Qfs (Music et Caya (2007)), a été exploitée dans le cadre de ce projet. Notons cependant que le modèle climatique fournit plusieurs autres variables qui pourraient trouver emploi dans les travaux futurs.

3.2.2 Couplage de modèles

Une telle utilisation de données climatiques dans une recherche portant sur le calage d'un modèle hydrologique fait appel à la notion de couplage de modèles. L'utilisation du couplage de modèles hydrologique et atmosphérique, principalement pour l'évaluation des impacts en contexte de changements climatiques, a connu son essort au début des années 2000 et a été mise à profit dans plusieurs études, notamment par Bindlish et Barros (2000), Robert *et al.* (2000) ainsi que Soulis *et al.* (2000). Deux types de couplage peuvent être distingués en fonction du degré d'interactions des modèles : les modèles dit liés (*one-way coupling*) et les modèles véritablement couplés (*two-ways coupling*). La première approche consiste à effectuer les étapes de modélisation de façon successive, sans interactions. Ainsi, le climat sera préalablement simulé, puis les sorties du modèle climatique serviront à alimenter le modèle hydrologique (Pietroniro *et al.* (2001)). Ce type de couplage est très courant, et a été utilisé par Carrer *et al.* (2009), notamment, afin d'estimer les effets des changements climatiques sur la recharge des bassins versants de l'Est du Canada, mais aussi par Mimikou *et al.* (2000) et Prudhomme *et al.* (2003). C'est aussi ce type de couplage qui prévaudra dans le cadre de cette recherche puisque les variables du modèle climatique seront utilisées afin de guider la paramétrisation du modèle hydrologique vers la physique (voir la Section 4.3.3 pour les détails de l'utilisation des données climatiques). Notons que le processus d'évapotranspiration est ainsi simulé de façon indépendante par le modèle climatique et le modèle hydrologique, et c'est cette particularité qui permettra l'application de la méthode développée dans ce mémoire. La deuxième stratégie de couplage de modèles implique un échange d'informations complet entre les deux types de modèles, ainsi qu'un schéma de surface commun, comme CLASS par exemple. Les processus hydrologiques sont donc simulés une seule fois par les deux modèles. Le couplage de WATCLASS au Modèle régional de climat canadien (MRCC) par le Centre canadien de la modélisation et l'analyse climatique (CCmaC) en est un exemple valant la peine d'être mentionné. Ce type de couplage est délicat à mettre en place pour plusieurs raisons, notamment puisqu'il demande un support informatique puissant, puisque les rivières ne sont pas modélisées par les modèles climatiques et puisqu'une conversion d'échelles est nécessaire (les bassins versants étant modélisés à une échelle inférieure à 1 km pour les modèles hydrologiques et à une échelle supérieure à 45 km pour les modèles climatiques). Ainsi, très peu d'exemples de ce type de couplage sont répertoriés dans la littérature.

3.3 Modélisation hydrologique

Comme le mentionne Gnouma (2006), afin de pouvoir étudier les problématiques liées à la gestion des ressources hydriques, pour dimensionner certains ouvrages ou pour encadrer la production d'hydroélectricité par exemple, la modélisation hydrologique devient vite un incontournable. Il n'est donc pas surprenant d'apprendre que plus d'une centaine de modèles hydrologiques différents existent, chacun avec ses caractéristiques propres : un nombre de paramètres plus ou moins grand, une modélisation de certains processus hydrologiques plutôt que d'autres, différents intrants et extrants, etc. Plusieurs études, entre autres Franchini (1991), Chiew (1993) et Gan *et al.* (1997), ont au fil des ans tenté de juger de la qualité de ces modèles afin de les classer et, éventuellement, d'en choisir un globalement meilleur que les autres. Malheureusement, les conclusions de ces études restent floues et difficiles à généraliser. En effet, plusieurs facteurs viennent complexifier l'analyse, notamment la disponibilité des données exigées en intrants (physiographiques et hydrométéorologiques), les besoin au niveau des extrants, la qualité de la modélisation des processus internes exigée, le type de bassin à modéliser, l'aisance de l'utilisateur avec un modèle, les délais impartis pour la modélisation, etc. Ce sont donc les besoins et les ressources des usagers qui permettent de départager les différents types de modèles.

3.3.1 Classification des modèles

Plusieurs classification ont été proposées, notamment par Singh (1995), Refsgaard et Storm (1996) et Chocat (1997), mais la plupart peuvent être résumées par celle présentée au Chapitre 8 du livre de Anctil *et al.* (2005). Ceux-ci proposent un schéma permettant de classifier les modèles hydrologiques selon leurs définitions du processus d'écoulement de l'eau, du domaine spatial et du domaine temporel. Ainsi, chaque modèle sera caractérisé par une combinaison de trois étiquettes.

Processus d'écoulement

La définition du processus d'écoulement se réfère à la structure mathématique interne du modèle et peut être subdivisée en deux sous-classes, soit les modèles offrant une modélisation dérivée des processus physiques et les modèles utilisant une modélisation dérivée des données. Dans le premier cas, chaque processus hydrologique devant être modélisé sera défini par une équation mathématique afin de respecter au mieux les mécanismes d'écoulement de l'eau. Pour un modèle dit conceptuel, ces équations seront simplifiées de façon plus

ou moins importante et ne seront pas nécessairement basées sur la physique, contrairement aux modèles dits physiques. Évidemment, de telles approximations au sein du modèle ont un prix en termes de qualité des solutions, mais le gain de temps de calcul vaut souvent la peine. De plus, l'utilisation d'un modèle physique demande une quantité de données souvent impossible à recueillir sur de grandes surfaces, notamment pour des questions de coûts ou de disponibilité des ressources. La deuxième sous-classe de modèles, dérivant des données sa modélisation des processus d'écoulement, pourrait être qualifiée d'agrégée, en ce sens où les processus hydrologiques ne sont pas définis explicitement. Ces modèles considèrent les intrants, souvent des précipitations, et tentent de formuler une équation les reliant aux extrants, généralement des débits, sans étapes intermédiaires, en utilisant des méthodes statistiques et les bases de données disponibles. Ces modèles dits empiriques, encore plus simplifiés que les modèles conceptuels, sont peu utilisés en pratique et servent principalement pour des études préliminaires ou pour des applications simples.

Domaine spatial

La catégorisation par modélisation du domaine spatial s'effectue via deux types de modèles : globaux et distribués. Un modèle global moyenne les observations, et parfois les caractéristiques physiographiques, du bassin versant afin de travailler avec une seule valeur pour l'ensemble la superficie étudiée. Le bassin est alors considéré comme une territoire ayant un comportement uniforme, tant au niveau du type et de l'occupation du sol que des conditions météorologiques. Par opposition, un modèle distribué est subdivisé en petites régions ayant un comportement et des caractéristiques semblables. Chacune possède ses propres intrants et fait circuler l'eau de manière à reproduire le comportement naturel du bassin versant. Les sorties du modèle sont donc aussi spatialement distribuées et peuvent être simulées pour n'importe quelle sous-section du bassin versant. Ces modèles sont donc beaucoup plus performants dans leur tentative de reproduction de la réalité hydrologique d'un bassin versant, mais sont en contrepartie très exigeants en termes de base de données et d'efforts d'implantation.

Domaine temporel

Finalement, la discrimination par définition du domaine temporel engendre deux types de modèles, soit événementiels et continus. Comme leur nom l'indique, les modèles dits événementiels servent principalement pour l'étude d'événements hydrologiques particuliers, souvent extrêmes, comme la crue maximale probable par exemple. Ce type de modèles nécessite de

pouvoir qualifier les conditions climatiques ayant précédé l'épisode météorologique sous investigation. Les modèles continus, quant à eux, simulent le comportement hydrologique d'un bassin versant à chaque pas de temps sur un horizon temporel plus étendu, soit de plusieurs années. De tels modèles sont utilisés lorsque l'ensemble du comportement d'un bassin versant est d'intérêt et permettent de suivre l'évolution continue d'un débit ou de variables d'état, et ce, tant que des intrants sont disponibles.

3.3.2 HSAMI, modèle hydrologique

Le modèle hydrologique utilisé dans le cadre de cette recherche, HSAMI (Service hydrométéorologique apports modulés intermédiaires), entre dans la catégorie des modèles de type conceptuel global continu. Dans l'optique où les modèles hydrologiques sont naturellement difficiles à optimiser, il semblait plus avisé de choisir un modèle nous permettant de réaliser plusieurs simulations en peu de temps. Dans le cas qui nous intéresse, une évaluation du modèle physique distribué continu Hydrotel, aussi utilisé par Hydro-Québec, prendrait une à deux heures alors que le modèle conceptuel HSAMI ne demande qu'une à deux secondes. Lorsqu'une méthode prometteuse sera identifiée, rien n'empêchera alors de transposer l'expérience sur un modèle plus complexe. Notons que le choix d'un modèle dérivant les processus à partir des données aurait été encore moins exigeant en termes de puissance de calcul, mais la modélisation des processus hydrologiques intermédiaires était primordiale à la mise en œuvre de l'approche exposée dans ce mémoire.

Introduction au modèle

HSAMI fut développé en 1983 par et pour Hydro-Québec, plus précisément par Bisson et Roberge (1983), puis fut raffiné par Fortin (1999). Depuis presque trente ans, ce modèle est utilisé comme appui dans la gestion des installations hydroélectriques par l'unité Prévisions et ressources hydriques d'Hydro-Québec. Plus récemment, il fut employé dans l'étude des changements climatiques, notamment par Minville *et al.* (2010), Laforce *et al.* (2011) et Roy (2011). Cet emploi fréquent est lié à sa capacité à reproduire la relation météo-apports le plus fidèlement possible à partir de données météorologiques telles que la température minimale et maximale ainsi que la quantité de précipitations, pluie et neige, pour une série de journées données. D'autres informations, facultatives, peuvent aussi être fournies en intrants au modèle et sont présentées en annexe, au Tableau B.1, tiré du mode d'emploi écrit par Fortin (1999).

Ces données météorologiques sont ensuite converties, remaniées et corrigées par les équations mathématiques supportant le modèle afin de simuler onze variables internes, ou processus hydrologiques, figurant en annexe, au Tableau B.4. Celles-ci peuvent être résumées en cinq catégories, soit l'évapotranspiration et la sublimation, l'accumulation et la fonte de la neige, les écoulements verticaux et horizontaux et, finalement, le gel et le dégel de l'eau contenue dans la neige ou le sol. Tout au long de ce processus, l'eau fournie en intrant au modèle est dirigée, à l'aide d'hydrogrammes unitaires et d'équations, à travers les trois réservoirs linéaires en cascades représentant les écoulements dans les sols en suivant le schéma conceptuel présenté à la Figure 3.1 afin d'arriver, au final, à une estimation journalière ou horaire, selon l'utilité, du débit ou apports naturels à l'exutoire du bassin versant étudié.

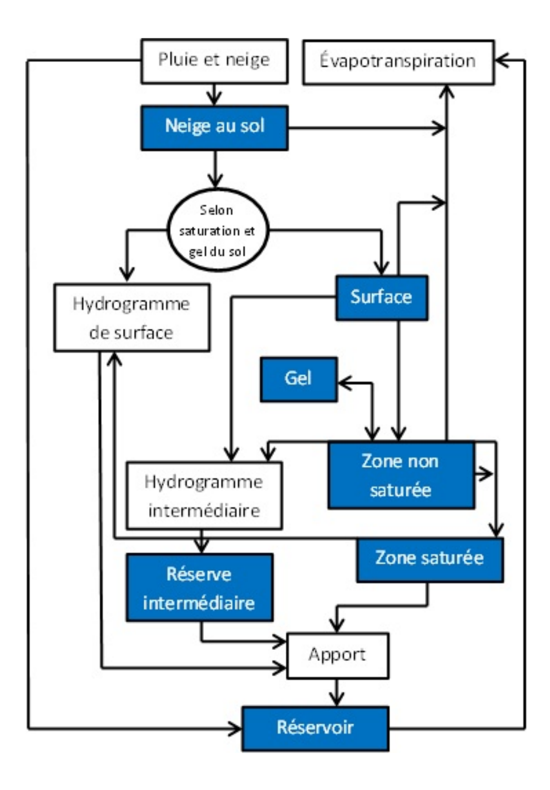


Figure 3.1 Schéma conceptuel du modèle HSAMI tiré de Fortin (1999).

Les paramètres

Vingt-trois paramètres, classés en cinq grandes familles, qui sont essentiellement les mêmes que celles classifiant les variables d'état, soutiennent ce modèle et apparaissent aux Tableaux B.2 et B.3. Ces paramètres ne peuvent être mesurés puisqu'ils intègrent le comportement et les caractéristiques moyennes d'un bassin. Ils doivent donc être ajustés par l'utilisateur, ou comme dans le cas présent, par un optimiseur, afin de guider le modèle

vers une représentation réaliste du comportement hydrologique d'un bassin versant. Puisque plusieurs paramètres ont un impact sur les mêmes processus hydrologiques, une relation complexe d'interdépendance relie certains d'entre eux. Ainsi, il n'est pas rare qu'un paramètre vienne en compenser un autre, et que les impacts d'un événement soient difficiles à cerner. Ce rapport entre les paramètres deviendra d'ailleurs un défi de taille lors de l'optimisation du modèle et sera repris lors de la discussion des résultats.

Les bornes

Les bornes utilisées pour encadrer les paramètres sont présentées au Tableau C.1 de l'Annexe C. Elles découlent d'une première version élaborée il y a plusieurs années à partir des paramètres de l'Outaouais utilisés par Hydro-Québec, et ont été ajustées et resserrées au fil du temps par plusieurs spécialistes suite aux résultats obtenus lors du calage de certains bassins versants. La version courante, établie en février 2003, est justifiée dans le rapport de Mathieu (2003). Une légère modification a toutefois été apportée à la borne supérieure du deuxième paramètre afin de fournir au modèle un espace paramétrique plus adapté à la reproduction du processus de l'évapotranspiration, et ce, dans des limites physiquement réalisables. Les bornes de l'Institut national de la recherche scientifique (INRS) (Annexe D) ont aussi été considérées, mais l'espace paramétrique engendré était beaucoup plus vaste, complexifiant grandement le problème, et l'inclusion de certaines zones paramétriques était contraire à la physique des processus, laissant place à de potentielles solutions aberrantes. Elles ont donc été rejettées pour cette application, mais feront l'objet d'expérimentations plus approfondies dans des travaux futurs.

Les données météorologiques

Le choix des données passées en intrants au modèle hydrologique découle d'une étude réalisée par Minville (2012) comparant deux ensembles de données météorologiques, celles produites à l'Institut de recherche d'Hydro-Québec (krigIREQ) et celles du service national d'information sur les terres et les eaux (*National Land and Water Information Service*) (NLWIS), à des sources externes sur la base des composantes du bilan hydrologique observé à long-terme. Les données krigIREQ ont été obtenues par krigeage, tandis que les données NLWIS proviennent d'une interpolation avec des splines effectuée par Environnement Canada. Jusqu'à maintenant, l'ensemble de données NLWIS était favorisé pour les recherches dû à sa disponibilité. Cependant, son application dans un projet d'Hydro-Québec a soulevé des doutes quant à sa capacité à bien représenter les bassins du Nord, qui revêtent une impor-

tance capitale vu leur fort potentiel hydro-électrique. L'ensemble de données krigIREQ a donc été envisagé, mais, n'ayant fait l'objet d'aucune étude analytique ou comparative, l'étude de Minville, citée plus haut, a dû précéder son utilisation. Deux points saillants peuvent être extraits de son rapport : en premier lieu, pour une utilisation dans le sud-ouest du Québec, les deux ensembles de données se sont révélés équivalents au niveau des bilans hydrologiques. Ensuite, pour les secteurs du nord et du nord-est, une quantité plus importante de précipitations a été générée par les données krigIREQ, ce qui a entraîné une modélisation du ruissellement et de l'évapotranspiration réelle plus réaliste. Des analyses complémentaires ont même conclu à une variabilité naturelle des précipitations analogue à celle réellement observée. En somme, en regard des conclusions de Minville (2012), les données météorologiques observées fournies comme intrant au modèle HSAMI seront les données météorologiques obtenues par krigeage à l'IREQ (krigIREQ).

Le site d'étude

Le site d'étude préliminaire, quand à lui, est le bassin versant Caniapiscau. Situé au sud du 54e parallèle de latitude Nord et à l'ouest du 70e méridien de longitude Ouest, il se trouve approximativement au centre du Québec. Avec une superficie de 37 328 km^2, ce bassin versant se trouve à être le bassin de tête du complexe La Grande, ce qui constitue la raison principale justifiant ce choix. Le principal avantage des bassins de tête est leur relative fiabilité au niveau de la simulation, puisque les erreurs potentielles sont cumulées tout au long du parcours de l'eau. En choisissant un bassin de tête donc, nous limitons au mieux ce type d'erreurs des données. Une autre raison consolidant ce choix est que les résultats de l'étalonnage de ce bassin provenant d'un autre projet étaient disponibles, ce qui nous offre une occasion de valider nos solutions sans coûts supplémentaires. Notons toutefois que le bassin versant Caniapiscau comporte un très grand réservoir et que le modèle HSAMI traite difficilement la pluie tombant directement dans un plan d'eau. Les résultats pourraient donc être influencés à la baisse par ce facteur.

CHAPITRE 4

DESCRIPTION DU PROBLÈME D'OPTIMISATION

Maintenant que le contexte, tant mathématiques qu'hydrologique, est mis en place, le problème visé par ce mémoire sera abordé. Cette section exposera le problème d'optimisation d'un point de vue mathématique, puis viendra détailler et justifier le choix des options de l'optimiseur. Finalement, les stratégies proposées afin de répondre aux besoins du calage en contexte de changements climatiques seront présentées.

4.1 Définition de la boîte noire

La boîte noire à laquelle sera couplé l'optimiseur NOMAD est une simulation informatique dont la pertinence sera évaluée par le moyen d'une fonction objectif. Le modèle soutenant la simulation, HSAMI, a été détaillé dans la Section 3.3.2. Cette section viendra plutôt exposer la boîte noire de plus haut niveau. En termes purement mathématiques donc, le problème d'optimisation dont il est question dans ce projet en est un de minimisation mono-objectif à 23 variables, appelées paramètres dans ce mémoire (voir l'introduction, Section 1). Trois contraintes sont déjà implantées dans le modèle et viennent régulariser les relations entre certains paramètres afin que leur valeur soit conforme à la réalité. Ces contraintes ne seront pas modifiées dans le cadre de cette recherche et sont regroupées sous une seule condition traitée par la méthode de la barrière extrême de par sa nature non relaxable.

4.1.1 La fonction objectif

L'objectif, soit le critère de Nash-Sutcliffe modifié ou $(1-ns)$, a été calculé en comparant les débits simulés par le modèle ainsi que les débits observés (apports hydrologiques obtenus par bilan hydrique) à l'exutoire du bassin versant Caniapiscau, une année sur deux pour les années 1961 à 2005. Hydrologiquement, les valeurs ainsi obtenues sont les valeurs en calibration. Le critère de Nash-Sutcliffe (défini à l'Équation 4.1), a fait l'objet de plusieurs critiques au fil des ans, notamment par Garrick *et al.* (1978), mais il reste malgré tout très utilisé par la communauté des hydrologues, comme le mentionne Scheafli et Gupta (2007).

$$ns = 1 - \frac{\sum_{t=1}^{T}(Q_o^t - Q_s^t)^2}{\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})^2}. \tag{4.1}$$

où les variables Q_o^t et Q_s^t font respectivement référence aux débits observés et simulés à chaque pas de temps t (allant du jour 1 au dernier jour (T) de la période de calibration), et où $\overline{Q_o}$ représente la valeur moyenne des débits observés sur toute la période de référence.

Cette fonction cherche donc à quantifier le gain obtenu par simulation par rapport à une utilisation de la simple moyenne des débits observés. Ainsi, dans le cas de la fonction $1 - ns$, une valeur se rapprochant de zéro serait associée à un modèle représentant parfaitement les données observées, tandis qu'un résultat proche de 1 indiquerait que le modèle ne donne pas de meilleurs résultats dans la simulation des débits qu'un autre se servant de la moyenne globale comme estimation journalière. Il peut aussi être pertinent de noter qu'une telle fonction a tendance à accorder plus de poids aux valeurs importantes, donc à la crue lorsqu'appliquée sur des débits.

Les années impaires, quant à elles, seront conservées afin de valider les prédictions du modèle calibré, ce qui constitue une pratique courante en hydrologie. Pour ce faire, les 23 paramètres optimaux obtenus après optimisation seront fixés et le modèle sera relancé avec les années impaires pour intrants. La performance du modèle en mode prédictif, ou en validation, sera ensuite mesurée, toujours avec le critère de Nash-Sutcliffe calculé sur les débits. La valeur ainsi obtenue est un indicateur de la performance espérée du modèle dans la prédiction de séries météorologiques différentes de celles utilisées pour paramétrer, et c'est pourquoi elle sera présentée dans la section présentant les résultats afin de compléter l'information fournie par la valeur obtenue en calibration.

La décision de procéder en mono-objectif n'est pas liée à une limitation de l'optimiseur, puisque NOMAD gère le multi-objectif, comme il a été mentionné dans la Section 2.1.2, mais bien à un choix de méthodologie. Comme Khu et Madsen (2005) le mentionnaient, avec l'avènement de l'optimisation multi-objectif, le problème de calage de modèle hydrologique en devient un de prise de décisions puisque le chercheur est amené à choisir le meilleur compromis, selon lui, entre les multiples objectifs et les nombreux compromis trouvés. C'est donc afin de conserver une démarche plus objective qu'un seul objectif sera utilisé. Par contre, rien n'empêche de reformuler les objectifs supplémentaires en contraintes, et c'est d'ailleurs une approche qui sera mise de l'avant dans de futurs travaux.

4.1.2 Lissage de la fonction

Afin de faciliter nos recherches, l'outil d'optimisation EAUPTIM, connectant deux modèles hydrologiques (Hydrotel et HSAMI) à deux algortihmes d'optimisation (MADS et NSGA-II de Deb *et al.* (2002)), a été scindé pour ne garder que les outils nécessaires à ce projet. À la suite des premiers tests réalisés sur la boîte noire, force était d'admettre que la performance de NOMAD était légèrement moins satisfaisante qu'avant la séparation d'avec NSGA. En sondant le code informatique reliant ces deux algorithmes, nous avons repéré que les instructions étaient telles qu'un arrondi était appliqué sur les paramètres avant d'être soumis à la boîte noire. En supposant que la fonction d'optimisation d'un modèle hydrologique est très bruitée, comme Duan *et al.* (1992) en font mention, l'arrondissement de la variable dépendante (un moins le critère de Nash-Sutcliffe) ou des variables indépendantes (les 23 paramètres du modèle) auraient pour effet de lisser la fonction, ce qui faciliterait grandement le travail d'un optimiseur utilisant la recherche directe. En effet, un tel lissage permet de diminuer de façon importante le nombre de minima locaux, ce qui limite le risque de converger vers un minimum local qui n'est pas un minimum global, ou de rester coincé dans une vallée peu prometteuse. Cet arrondi permet aussi un filtrage du bruit numérique.

L'hypothèse valait la peine d'être explorée, puisque la pertinence d'une telle précision sur les données simulées avait déjà été remise en question. En effet, les mesures réelles (observées) sont loin de pouvoir atteindre ne serait-ce que la moitié d'un tel degré d'exactitude, il est donc inutile d'en exiger autant des données simulées. De plus, les différences entre deux valeurs de Nash-Sutcliffe ne sont plus considérées comme étant significatives passé un certain seuil. Ainsi, après avoir analysé la question sous plusieurs angles, l'application de l'arrondi a été intégré directement dans la boîte noire, et sur la variable dépendante seulement. La précision des paramètres, ou variables indépendantes, a quant à elle été traitée par un critère d'arrêt qui sera détaillé plus loin. Cette façon de faire a été privilégiée notamment parce qu'elle répondait le mieux en termes de résultats, mais aussi parce qu'elle était plus naturelle à mettre en place. Malgré tout, le résultat final est le même, soit un lissage de la fonction par rapport à la variable dépendante, mais à l'intérieur de la boîte noire plutôt qu'à l'extérieur. Suite à l'application de cette procédure, non seulement les résultats étaient-ils significativement meilleurs [1], mais

1. Le test statistique de Kolmogorov-Smirnov a été appliqué de façon successive sur les résultats des trois stratégies de calage, en comparant chaque fois les résultats de cinquante optimisations réalisées avant et après le lissage de la fonction, afin de déterminer si les deux échantillons suivaient la même loi, ou s'ils pouvaient être considérés comme statistiquement différents. Les tests sont venus confirmer que chacune des paires d'échantillons était différente avec des p-value de $4.9988e^{-19}$, $4.9988e^{-19}$ et $4.0089e^{-15}$ pour les stratégies de calage traditionnel, de calage traditionnel avec bilan sur l'ETR et de calage parallèle respectivement. Ainsi, les résultats obtenus avec lissage sont non seulement statistiquement meilleurs, mais ont été obtenus plus rapidement et avec moins d'efforts de la part de l'optimiseur.

ils furent obtenus en moyenne en 7 000 évaluations plutôt qu'en 20 000, le tout en réduisant le temps de calcul de 80 %. De telles économies étant non négligeables, le concept de lissage de la fonction pourra être employé pour de futurs travaux, indépendamment des résultats liés aux méthodes développées dans ce rapport.

4.2 Conditions d'optimisation

4.2.1 Les options de NOMAD

Certaines options de NOMAD ont été fixées afin d'utiliser le plein potentiel de l'optimiseur pour chaque calibration. Après avoir effectué certains tests comparatifs, les fonctions optionnelles TMP_DIR, ASYNCHRONOUS, SNAP_TO_BOUNDS et VNS_SEARCH se sont révélées plus pertinentes lorsqu'elles n'étaient pas utilisées selon les valeurs par défaut et ont donc été mises à profit pour l'ensemble du projet. Chacune de ces fonctions ayant son importance et ses caractéristiques, elles seront passées brièvement en revue dans les sections suivantes.

TMP_DIR

La première, et peut-être la plus simple, est l'option TMP_DIR. Celle-ci n'a pas d'impact sur la solution optimale ni sur la méthode de résolution. Elle permet simplement d'accélérer le processus d'optimisation en mettant à profit le disque dur local, limitant ainsi certaines opérations informatiques (écritures et lectures) plus coûteuses en termes de temps de calcul.

ASYNCHRONOUS

La fonctionnalité suivante, ASYNCHRONOUS, est aussi conçue afin d'optimiser l'utilisation de multiples processeurs informatique. Son impact sur les résultats obtenus à la suite d'une calibration est cependant beaucoup plus important. L'activation de cette fonction implique que l'optimisation se déroule en parallèle d'un point de vue informatique, ce qui signifie que certaines tâches se déroulent simultanément sur différents processeurs. La façon dont la division des tâches est gérée par NOMAD a été précédemment discutée dans la section 2.1.2. Une recherche dite asynchrone encourage l'optimiseur à poursuivre son travail sur les évaluations reçues, sans attendre la réponse des autres processeurs, ce qui cause une variabilité entre les solutions. En effet, selon le degré d'activité de chaque processeur, possiblement sollicité par un ou plusieurs utilisateurs du réseau, avec un traffic fluctuant de façon aléatoire au fil du temps, chaque tâche sera plus ou moins longue à exécuter. Ainsi, un processeur détenant une solution améliorant de peu la solution courante pourrait terminer plus tôt qu'un proces-

seur ayant trouvé un point encore meilleur, auquel cas ce point ne sera pas nécessairement retenu pour une exploration plus approfondie. Le chemin parcouru dans l'espace des solutions devient donc dépendant de la vitesse des processeurs, ce qui cause une variabilité entre les solutions optimales de deux calibrations en tous points identiques par ailleurs.

Certains autres facteurs rendant les solutions difficiles, sinon impossibles, à reproduire de façon exacte doivent aussi être mentionnés. Selon Beven (1996), le problème d'optimisation d'un modèle hydrologique n'a généralement pas de solution unique. Cette notion, l'équifinalité, est basée sur le principe qu'aucun modèle hydrologique, ni aucun ensemble de paramètres ne peut être considéré comme parfaitement représentatif de la réalité. Cela est dû notamment à la non-linéarité du problème, aux nombreuses interactions entre les paramètres, à l'insensibilité de certains paramètres qui ne sont sollicités que lors d'événements extrêmes par exemple, aux observations parfois erronées servant à forcer le modèle et aux multiples simplifications de la physique soutenant les équations de la simulation. En conséquent, plusieurs solutions sont évaluées comme étant équivalentes, et ainsi, il est impossible d'arriver à une solution unique, ou du moins à une solution significativement supérieure à toutes les autres. La recommandation de Beven consiste à ne pas considérer seulement l'optimum, mais bien d'évaluer les solutions équivalentes dans leur capacité à reproduire de façon cohérente les processus physiques liés au scénario étudié. D'ailleurs, à la page 293 de son article, Beven (1996) affirme qu'une validation sur les processus internes du modèle plutôt que sur les années impaires serait beaucoup plus robuste. Cependant, toujours selon lui, un tel test de confirmation ne pourrait pas être mis en place, les procédés ne pouvant être mesurés adéquatement. Malgré le fait que ces hypothèses ne soient pas acceptées par l'ensemble de la communauté scientifique et que la recherche d'un optimum unique ait dirigé la conception de la méthodologie de ce projet, les propos défendus par Beven seront tout de même indirectement intégrés à la méthode développée dans ce rapport puiqu'elle viendra confronter les solutions sur leur capacité à modéliser correctement la physique des processus.

VNS_SEARCH

Un autre aspect compromettant la reproductibilité des résultats concerne la recherche aléatoire liée à l'utilisation de la VNS. Cette option, VNS_SEARCH, est détaillée dans la section 2.1.2 et a été ajustée spécifiquement pour chaque stratégie de calage. Mentionnons seulement que le point central de cette méthode est l'induction d'une série de descente dans l'espace des variables, mais cette recherche par bonds peut parfois amener l'optimiseur à explorer des zones de l'espace-solution très isolées qui auraient été impossibles à trouver seulement avec une recherche locale, et très improbable à retrouver avec l'aide de la méta-

heuristique [2].

SNAP_TO_BOUNDS

La dernière option, normalement activée par défaut, se nomme SNAP_TO_BOUNDS et fonctionne de la façon suivante : lorsqu'une ou plusieurs coordonnées d'un point à évaluer est située hors des bornes définies au préalable et fournies à l'optimiseur, NOMAD soumettra le point à la simulation en projettant la coordonnée problématique à la borne la plus près. En désactivant cette fonctionnalité, les points hors bornes seront rejetés avant même d'être évalués et ne seront par conséquent pas comptabilisé dans le nombre d'évaluations de la boîte noire. Cette façon de procéder permettra à l'optimiseur de se rapprocher plus finement de la borne sans être coincé directement sur celle-ci.

Choix du point de départ et du nombre de processeurs

Mis à part le choix des fonctionnalités de l'optimiseur, certaines décisions ont dû être prises concernant notamment le choix d'un point de départ pour l'algorithme MADS ainsi que le nombre de processus utilisés en parallèle. Nos hypothèses concernant le point de départ était qu'en exécutant un grand nombre d'évaluations et en activant l'option VNS_SEARCH, le point de départ tendait à perdre de son importance. Nous avons tout de même validé ces hypothèses avant de commencer et la corrélation entre cinquante points de départ choisis aléatoirement et leur point d'arrivée est presque nulle. C'est pourquoi, pour chacune de nos optimisations, le point de départ a été choisi de façon aléatoire.

La question du nombre de processeurs est un peu plus complexe. Tel que mentionné à la Section 2.1.2, l'algorithme MADS effectue une exploration de l'espace en suivant certaines directions orthogonales entre elles. Le nombre de directions différentes suivies à chaque itération est fonction du nombre de variables du problème à optimiser, par défaut celui-ci est le double du nombre de variables. Dans le cas présent, le modèle HSAMI possède vingt-trois paramètres à ajuster, ce qui conduit à 46 directions visitées à chaque itération. En choisissant plus de 46 processeurs, certains ne seraient pas mis à profit, ce qui fait que les options se limitent donc à l'intervalle [1 ;46]. De là, un nombre de processeurs trop faible viendrait

2. Par exemple, une solution présentant une excellente valeur de l'objectif a été trouvée durant cette recherche et n'a jamais pu être égalée. Ce point a donc été investigué afin de comprendre pourquoi il ne pouvait plus être retrouvé. La raison est liée à la forme de la fonction à cet endroit, qui pourrait être comparée à un volcan. Ainsi, en se rapprochant du point d'intérêt, la fonction objectif est de plus en plus dégradée pour ensuite redescendre brutalement sur une toute petite partie de l'espace. À moins de tomber directement sur l'excellente solution, un algorithme classera cette zone de l'espace comme n'étant pas prometteuse et ne l'explorera pas en détails.

annuler le gain de performance lié à l'utilisation du parallélisme, et un nombre trop élevé contrecarrerait l'option ASYNCHRONOUS, puisque tous les points seraient visités simultanément, ce qui occasionnerait aussi une augmentation du temps de calcul sans fournir de gains proportionnels au niveau de l'objectif. De ces constatations, plusieurs tests ont été menés sur les possibilités restantes afin de conclure à de très faibles variations de l'objectif et du temps de calcul. Un nombre de huit processeurs a finalement été fixé pour toutes les optimisations réalisées dans le cadre de ce projet.

Il est important de noter que les choix mentionnés dans cette section ne sont pas valables pour tous les problèmes d'optimisation résolus à l'aide de NOMAD, ni même pour le calage de modèles hydrologiques par NOMAD, mais bien les options les plus judicieuses pour la calibration du modèle HSAMI, dans les conditions de la présente étude. Pour plus d'information sur les options disponibles et sur leur utilisation, le lecteur est encouragé à consulter le manuel d'utilisation de NOMAD (Le Digabel et Tribes (2009)).

4.2.2 Critères d'arrêt

L'optimiseur NOMAD considère plusieurs critères provoquant l'arrêt du processus d'optimisation. La première condition rencontrée parmi celles proposées par l'utilisateur viendra terminer l'itération en cours et la solution courante sera considérée comme l'approximation de la solution optimale. Dans le cadre de cette recherche, trois critères ont été appliqués, mais plusieurs autres sont disponibles, comme MAX_TIME, spécifiant une durée maximale pour l'exécution de l'algorithme, STOP_IF_FEASIBLE qui clôt la séquence d'optimisation dès qu'une solution réalisable est trouvée ou F_TARGET, qui interrompt NOMAD quand la solution courante est inférieure ou égale à une valeur spécifiée à priori, pour ne nommer que ceux-là. Ce sont les besoins de l'utilisateur relativement aux particularités de son problème qui guideront ces choix.

Le critère d'arrêt s'étant révélé le plus pertinent pour le problème actuel, MIN_MESH_SIZE, constitue une façon détournée de limiter la précision sur les variables. Des tests préliminaires ont montré qu'une plus grande précision sur les paramètres ne se traduisait pas en de meilleurs résultats au niveau de l'objectif. Une explication de ce phénomène a été avancée par M. Minville suite à une étude réalisée en 2011 ayant relevé que le comportement hydrologique du modèle devenait linéaire lorsque la fonction était examinée sur des sous-sections réduites du domaine paramétrique. Ainsi, il a été possible de conclure que la réaction de la fonction objectif à une modification inférieure ou égale à 6.25% de la valeur d'un paramètre pouvait

être estimée par une approximation linéaire, et donc qu'aucune fluctuation importante ne pouvait être espérée. Par contre, l'exploration à de telles échelles devient vite coûteuse en termes de temps de calcul puisque la grande majorité des évaluations prennent place au cœur de la vallée prometteuse, où peu de gains sur l'objectif sont constatés, comme le démontre la Figure 4.1. Celle-ci exprime l'évolution de la valeur de la fonction objectif en fonction du nombre d'évaluations de la boîte noire, et montre clairement la ténuité de la zone d'intérêt. MIN_MESH_SIZE a donc été mis à profit comme critère d'arrêt afin de terminer l'optimisation si un des paramètres venait à être trop précis. L'intérêt de cette option de NOMAD est qu'elle permet à l'utilisateur de spécifier une précision personnalisée pour chaque variable, en fournissant à l'optimiseur un pourcentage de l'écart entre les bornes relatives à chaque paramètre. Ainsi, dans le cas présent, un écart de 0.00001 s'est révélé idéal en n'étant ni trop relâché, n'ayant pas l'impact souhaité sur le nombre d'évaluations, ni trop serré, interrompant la séquence d'optimisation prématurément.

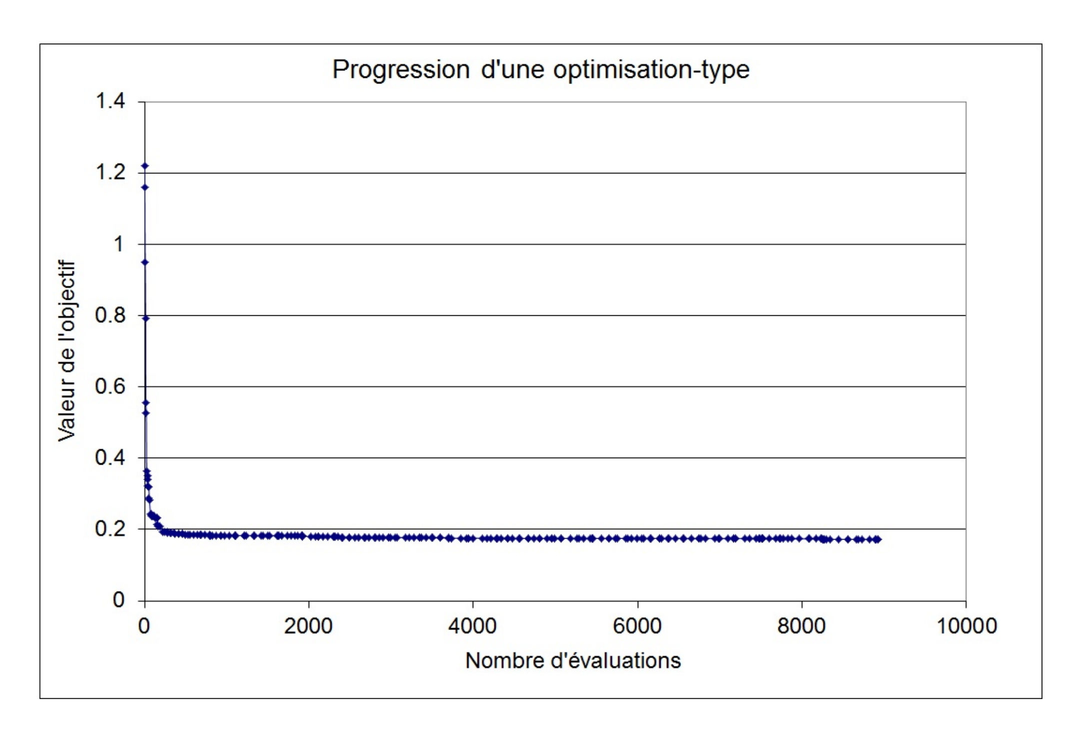


Figure 4.1 Évolution de la valeur de l'objectif en fonction du nombre d'évaluations pour une optimisation type.

Le deuxième critère d'arrêt spécifié, MAX_BB_EVAL, correspond au nombre maximal d'évaluations de la boîte noire. Ce paramètre a été fixé à 20 000 évaluations pour notre problème, un seuil qui était atteint dans environ 10 % des cas auparavant. Une conséquence du lissage de la fonction mentionné plus haut a cependant été de réduire grandement le nombre

d'évaluations de la fonction objectif ce qui fait que cette condition d'arrêt est devenue un filet de sécurité plutôt qu'un critère actif. Finalement, MIN_POLL_SIZE est régi par un paramètre par défaut n'ayant pas été modifié et détermine une taille minimale de zone de recherche en-deçà de laquelle l'amélioration de l'objectif n'est plus significative.

4.3 Les stratégies de calage

L'essence de ce projet repose sur trois stratégies de calage qui seront déployées afin de comparer leur robustesse ainsi que leur stabilité en mode prédictif. De ces trois stratégies, la première agira à titre de stratégie de contrôle, alors que les deux autres se veulent des avenues de recherche afin d'améliorer le rendement des modèles hydrologiques en contexte de changements climatiques. Ces trois méthodes sont représentées schématiquement sur la Figure 4.2 et seront détaillées une à une dans les prochaines sections.

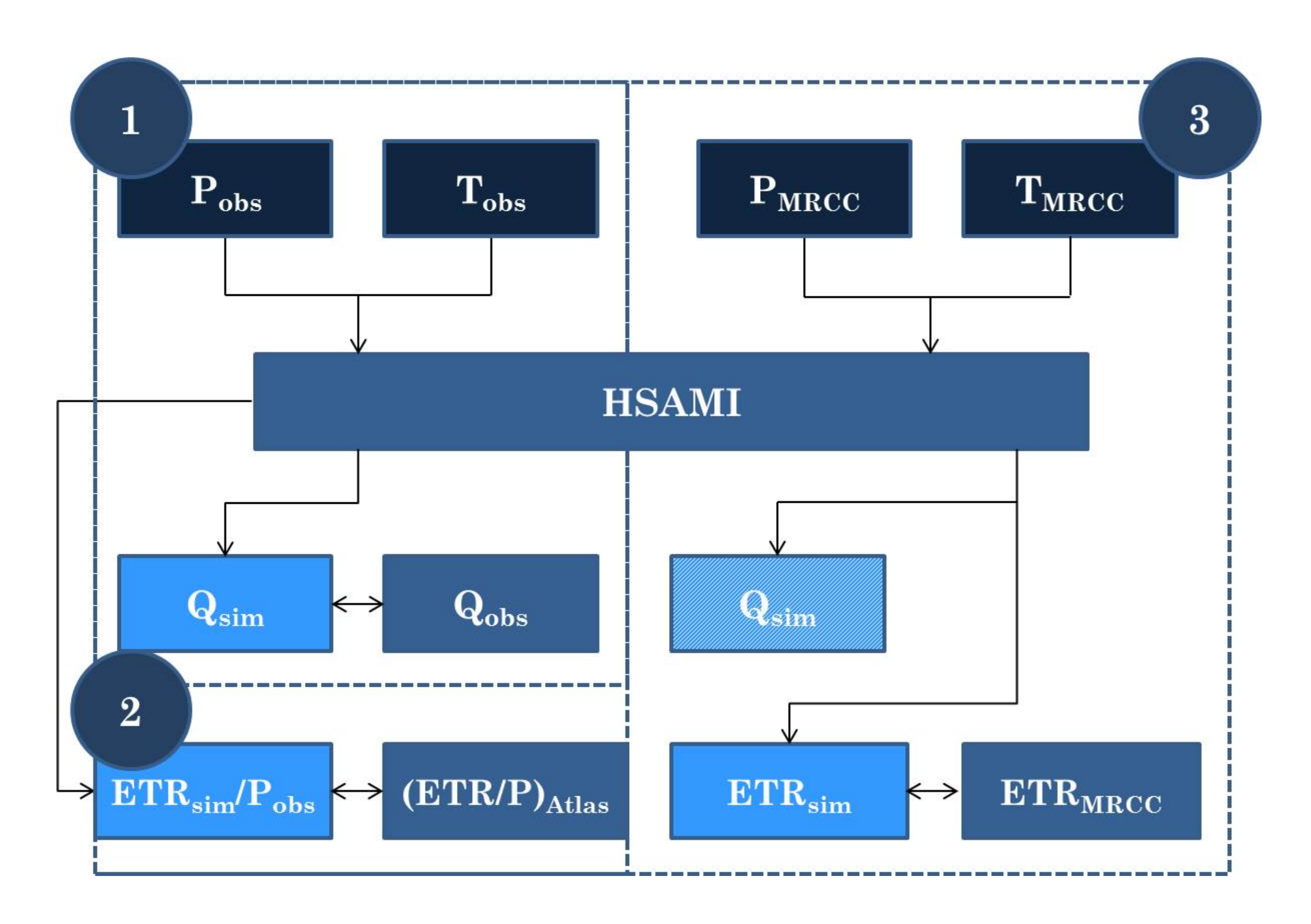


Figure 4.2 Schéma des trois stratégies de calage.

4.3.1 Calage traditionnel

La première, ci-après nommée calage traditionnel sans contrainte ou simplement calage traditionnel, consiste à optimiser le modèle hydrologique tel que présenté dans les Sections 3 et 4, sans qu'aucune nouvelle contrainte n'ait été ajoutée. La contrainte régissant les relations interparamètres du modèle reste, et restera, partie prenante du modèle. Cette stratégie servira de base de comparaison avec l'utilisation actuelle du modèle HSAMI chez Hydro-Québec. Elle est schématisée de façon individuelle à la Figure 4.3.

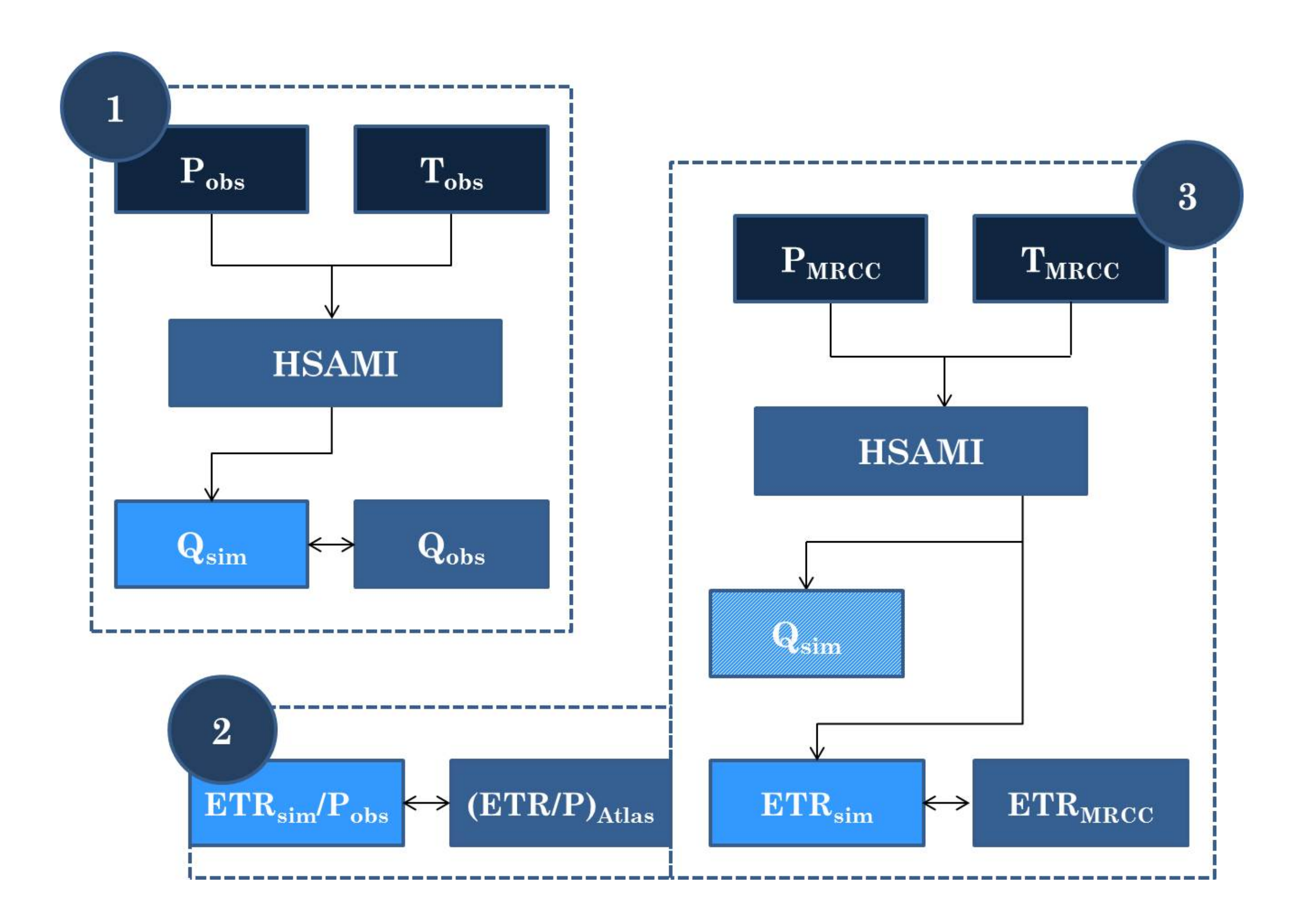


Figure 4.3 Schéma de la première stratégie de calage.

4.3.2 Calage traditionnel avec bilan sur l'ETR

La deuxième stratégie, nommée calage traditionnel avec contrainte ou calage traditionnel avec bilan sur l'ETR, consiste à lancer le problème de base en lui ajoutant une contrainte externe à la boîte noire (Figure 4.4). Cette contrainte vise à diriger une variable d'état du

modèle, soit l'évapotranspiration. La série journalière simulée par le modèle sera cumulée puis divisée par la somme des précipitations observées ($\frac{ETR}{P}$), et cette valeur sera comparée à une valeur théorique tirée de l'Atlas du Canada (voir Annexe E). Pour le bassin versant Caniapiscau, la valeur interpolée est de 25 %. Considérant la possible obsolescence de l'Atlas du Canada, ayant été produit en 1978, et l'erreur induite par la méthode d'interpolation, exiger du modèle la valeur théorique exacte ne semblait pas approprié. Afin de décider d'une zone de valeurs acceptables, le même calcul ($\frac{ETR}{P}$) a été effectué, cette fois à l'aide des données du modèle climatique uniquement, pour une valeur de 29 %, ainsi qu'avec les données observées, avec un résultat de 22 %. En estimant que la valeur optimale devrait se trouver dans l'intervalle englobé par ces valeurs, la valeur de l'Atlas a été bornée par un écart de 4 %, ce qui a mené à la contrainte :

$$21~\% \leq \frac{ETR}{P} \leq 29~\%. \tag{4.2}$$

Cet intervalle de plus ou moins 4 % a été fixé de façon plutôt arbitraire en fonction des données mentionnées plus haut, mais une méthode plus objective sera développée dans de futurs travaux. L'objectif visé par l'ajout de cette contrainte est de forcer la dynamique interne du modèle vers une représentation plus physique de ce processus hydrologique, ce qui devrait améliorer la fiabilité en mode prédictif. L'ajout de contraintes en général a aussi un deuxième effet positif sur l'optimisation, énoncé par Cooper *et al.* (2007). Le calage d'un modèle hydrologique est complexe puisqu'une quantité importante de minima locaux est présente dans chacune des nombreuses dimensions du problème (vingt-trois dimensions dans le cas présent). En ajoutant une ou plusieurs contraintes, certaines zones de l'espace de recherche sont exclues, ce qui peut améliorer la performance de l'algorithme d'optimisation en éliminant plusieurs minima locaux et en réduisant l'espace paramétrique. Ainsi, la recherche est approfondie dans les zones plus prometteuses, puisque moins d'évaluations sont utilisées pour explorer un espace infiniment vaste.

4.3.3 Calage parallèle avec critère de Nash-Sutcliffe sur l'ETR

Finalement, la troisième stratégie (Figure 4.5), le calage parallèle avec contrainte, est beaucoup plus complexe. Le modèle hydrologique sera lancé deux fois simultanément : la première version sera identique au problème de base décrit par la première stratégie, tandis que la deuxième version servira à contraindre la première. Ainsi, à chaque évaluation, un même ensemble de paramètres sera utilisé afin de lancer les deux simulations. La première,

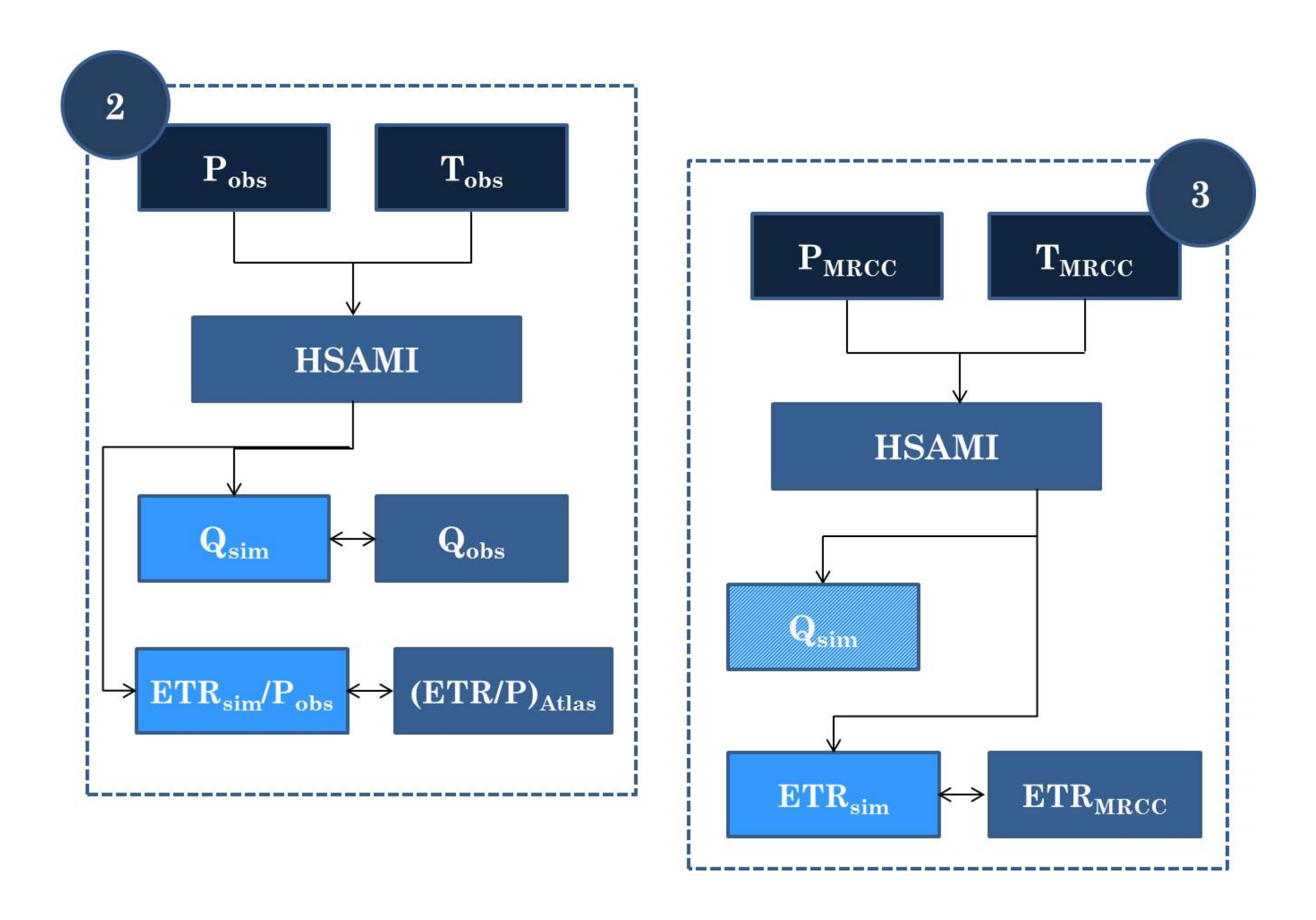


Figure 4.4 Schéma de la deuxième stratégie de calage.

ayant pour intrants les valeurs de précipitations et de températures observées de 1961 à 2005, servira à calculer la fonction objectif. La deuxième, quant à elle, recevra pour intrants les données météorologiques d'un modèle climatique, allant de 1961 à 2001, afin de simuler une série de valeurs d'évapotranspiration. Ces valeurs seront ensuite comparées à l'évapotranspiration produite par le modèle climatique pour ces mêmes années à l'aide du critère de Nash-Sutcliffe.

Validité de la comparaison

Nous pouvons considérer que le processus de l'évapotranspiration tel que simulé par le modèle climatique est équivalent à la variable d'état du modèle HSAMI. L'évapotranspiration totale, ou Qfs, du modèle climatique est constituée de cinq processus hydrologiques, soit l'évaporation de la pluie interceptée par la végétation (QFCL), l'évaporation de la pluie interceptée par la neige (QFCF), l'évaporation du sol nu (QFG), la sublimation et l'évaporation de l'eau contenue dans la neige (QFN) et finalement la transpiration de la végétation

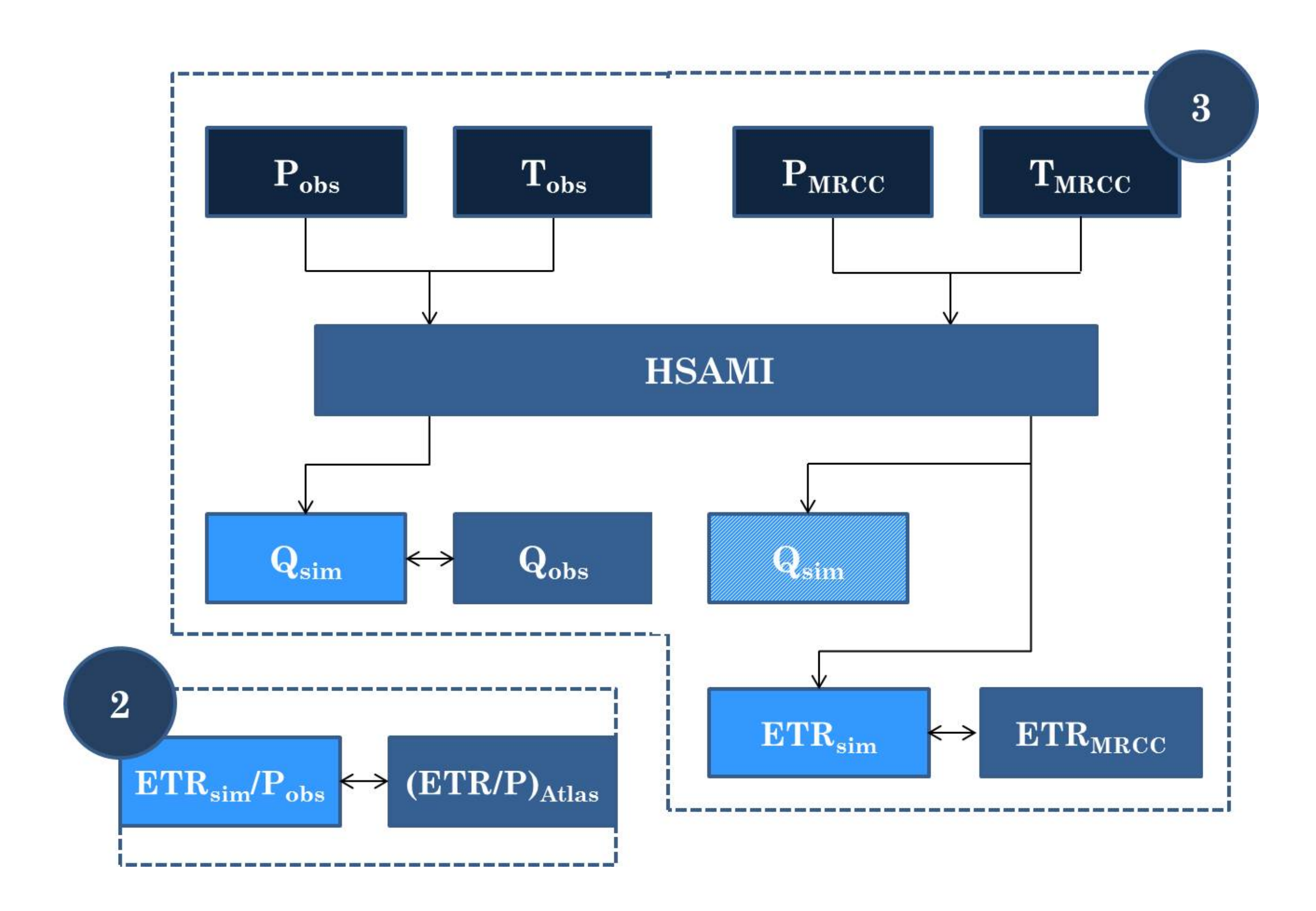


Figure 4.5 Schéma de la troisième stratégie de calage.

(QFC). HSAMI, quant à lui, simule quatre processus, qui sont grosso modo équivalent à ceux du MRC. Ces processus sont l'évaporation de la pluie interceptée par la végétation, la sublimation de l'eau contenue dans la neige, l'évaporation de l'eau contenue dans la neige et la transpiration de la végétation. De plus, Lofgren et al. (2011) ont déterminé que la corrélation entre la tempétaure de l'air et l'évapotranspiration était réduite au cycle annuel moyen de ces variables. Dans ce sens, puisque HSAMI se fonde sur les températures pour calculer l'évapotranspiration réelle, et que l'ETR de CRCM-ERA40 est fondée sur les bilans énergétiques, une relation existe entre ces trois variables et il est admissible de les mettre en relation.

En considérant que le modèle climatique, de par sa complexité, reproduit certains phénomènes physiques plus fidèlement que le modèle hydrologique, l'objectif poursuivi par l'ajout de cette dernière stratégie est de forcer le comportement interne du modèle hydrologique vers celui du modèle climatique. Cela devrait avoir pour effet d'améliorer la robustesse et la stabilité des prévisions, tout en facilitant les conditions d'optimisation grâce aux avantages liés à

l'ajout de contraintes mentionnés dans la section précédente. Malgré tout, il serait bon de noter que les contraintes du calage parallèle seront beaucoup plus délicates à mettre en place que celles du calage traditionnel sous contrainte. En effet, plutôt que de s'appuyer sur une valeur théorique, ces contraintes feront appel au jugement de l'utilisateur par le biais d'une mesure d'erreur. Tel que mentionné dans notre hypothèse de départ, une conséquence vraisemblable de l'ajout de contraintes sur la physique des processus consistera en une dégradation de la fonction objectif plus ou moins importante et c'est à l'hydrologue que reviendra le rôle de juger de la qualité de données de débits et de précipitations, afin de mettre le poids sur les données qu'il juge les plus fiables. Ainsi, pour certains bassins versants, l'objectif pourrait être dégradé de façon plus importante, d'autant plus si la qualité des débits observés est mise en doute. Une telle approche permet donc de bénéficier des source de données disponibles et de leur qualité présumée afin d'obtenir un calage le plus réaliste possible.

CHAPITRE 5

RÉSULTATS ET ANALYSES

Cette section est la pierre angulaire de toute cette recherche. L'ensemble des résultats obtenus au cours de ce projet sera présenté et les grandes conclusions quant à la validité de l'approche par contraintes seront tirées. Le but premier de ce projet étant de parvenir à suggérer une approche robuste, transposable en contexte de changements climatiques, les trois stratégies de calage seront comparées entre elles dans cette perspective. Au terme de ce chapitre, une stratégie, s'il en est une, pourra être proposée pour le calage du modèle HSAMI en contexte futur.

5.1 Analyses préalables

Avant d'obtenir les résultats permettant de comparer les stratégies de calage entre elles, il a été nécessaire de passer par une étape intermédiaire afin de fixer le membre de droite de la contrainte de la troisième stratégie, ainsi que le degré d'utilisation optimal de la VNS. Par la suite, toutes les stratégies de calage seront considérées comme étant exploitées à leur plein potentiel, et elles pourront être confrontées sur les plans de la variabilité et de la robustesse, ainsi qu'au niveau de leur performance du point de vue hydrologique.

5.1.1 Analyses de sensibilité

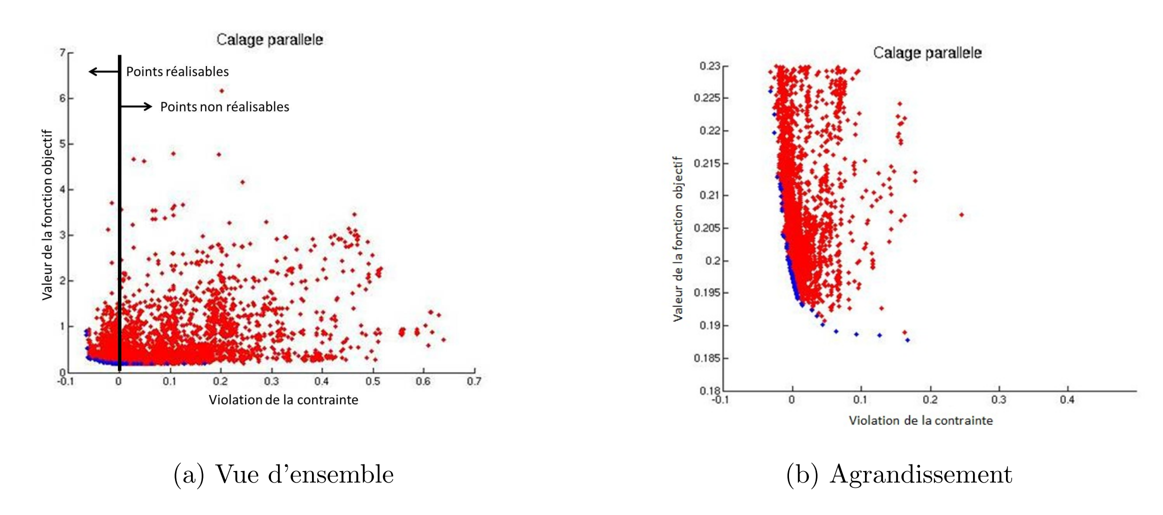
Telles que présentées à la Section 2.2.3, les analyses de sensibilité sont particulièrement utiles lorsqu'une contrainte peut être ajustée par l'utilisateur, afin de tirer profit de l'information disponible dans le but d'améliorer la formulation du problème. Dans le cas qui nous intéresse, seule la troisième stratégie de calage, soit le calage en parallèle, se prêtait à ce genre d'analyse. En effet, la première stratégie, le calage traditionnel, n'implique pas de contrainte, et la deuxième, le calage traditionnel avec bilan ETR, est un cas particulier. La contrainte de bilan d'ETR devrait être considérée comme non relaxable. En effet, l'hypothèse soutenant cette méthode est que l'Atlas du Canada fournit des bilans représentatifs des observations. Dans cette optique, une solution qui s'écarte des bornes n'a pas de valeur, car cela prouverait que la modélisation des processus internes est erronée. De ce fait, même si elle arrivait à bien représenter le débit, elle ne serait pas retenue pour le calage en contexte de changements climatiques. Ainsi, cette contrainte aurait normalement dû être gérée par la méthode de la barrière extrême, mais les chercheurs de l'Institut de recherche d'Hydro-Québec ont choisi de

traiter cette contrainte à l'aide de la barrière progressive afin d'étudier le comportement de leur modèle. Les analyses de sensibilité ont donc été réalisées sur ladite contrainte, mais les conclusions sortent du cadre de ce projet et en aucun cas la contrainte n'a été ajustée suite à ces analyses.

En premier lieu, une première optimisation du modèle contraint a été effectuée afin d'obtenir les données nécessaires à la réalisation d'une analyse de sensibilité à l'aide de l'outil d'analyse simple de NOMAD. La stratégie de calage parallèle a été sommairement expliquée à la Section 4.3.3, mais il pourrait être utile de rappeler que la contrainte du calage parallèle consiste à comparer sur une base journalière l'évapotranspiration réelle produite par le modèle HSAMI à celle fournie par le modèle climatique MRCC alimenté par *aft* à l'aide du critère de Nash-Sutcliffe. La contrainte de départ a été fixée comme étant un moins le critère de Nash-Sutcliffe (1-ns) inférieur à 0.4. Une valeur de 0.6 (ou de $1 - 0.4$) pour le critère de Nash-Sutcliffe est globalement acceptée par les hydrologues comme étant un résultat allant de raisonnable à bon comme le confirment plusieurs articles, notamment Chiew *et al.* (2009), Kralisch *et al.* (2007) et Pappenberger et Buizza (2009). C'est donc pour cette raison qu'une telle valeur a été choisie comme valeur initiale. Elle sera évidemment raffinée à l'aide d'analyses de sensibilité afin de vraiment tirer profit de la qualité et de la forme des données de notre bassin versant. Une première analyse de la contrainte, donc, est présentée à la Figure 5.1a.

Ce graphique, ainsi que les suivants, peut être lu de la façon suivante : tous les points représentent des évaluations effectuées préalablement par NOMAD dans le cadre de l'optimisation. Les points bleus dominent les rouges au sens Pareto pour un problème bi-objectif de minimisation de l'objectif initial $(1 - ns)$ ainsi que de la violation de la contrainte. L'abscisse du graphique représente donc la violation de la contrainte étudiée, soit, dans le cas présent, 0.4 diminué du résultat obtenu. Une valeur négative sera donc associée à un point respectant la contrainte. L'ordonnée du graphique, quant à elle, représente la valeur de l'objectif. Afin de bien cerner les variations en abcisse et en ordonnée, il sera préférable d'agrandir la zone d'intérêt des variables à l'étude et c'est pourquoi, à partir de ce point, seules les agrandissements seront présentés, à moins d'indications contraires. Ainsi, la figure suivante, 5.1b, permet de statuer sur la contrainte.

On peut voir qu'une importante amélioration de la fonction objectif est obtenue à la limite de l'activation de la contrainte, l'objectif passant rapidement de 0.23 à un peu moins de 0.195. Puisqu'une valeur de 0.6 pour le critère de Nash-Sutcliffe (ou de 0.4 pour $1 - ns$) était



(a) Vue d'ensemble

(b) Agrandissement

Figure 5.1 Analyse de sensibilité pour la contrainte $1 - ns(journalier) \leq 0.4$.

le plus bas que l'on était prêt à accepter pour ce bassin versant, et puisque l'amélioration de l'objectif au-delà n'est pas fulgurante, la contrainte a été fixée à ce point et 50 optimisations indépendantes ont été lancées afin d'avoir un échantillon représentatif. D'un point de vue mathématique, cela semblait convenable, mais c'est ici que le partage des connaissances entre les deux expertises, mathématique et hydrologique, prend toute son importance. En observant le cycle annuel de l'évapotranspiration réelle (présenté à la Figure 5.2), ou évapogramme, il était clair que les périodes incluant les jours juliens 30 à 150 et les jours 260 à 310 n'étaient pas bien modélisées, et donc que la définition de la contrainte ne représentait pas bien le problème. Cela n'est pas surprenant, car il est bien connu que le critère de Nash-Sutcliffe tend à mettre plus d'importance sur les forts débits, et donc sur la période de la crue lorsque les débits sont considérés, ou, dans le cas présent, sur la période estivale présentant un régime d'évapotranspiration plus important. La contrainte a donc été modifiée afin d'aller chercher toute l'information contenue dans le cycle annuel d'évapotranspiration réelle. Plusieurs méthodes ont été testées avant d'arriver à la stratégie finale, et elles seront survolées brièvement, afin de s'attarder comme il se doit à celle qui semblait la plus prometteuse. À chaque étape, les analyses de sensibilité nécessaires ont été effectuées, mais ne seront pas présentées pour ne pas alourdir le texte.

Ainsi, le premier test à avoir été réalisé fut de modifier la fonction objectif pour utiliser le PBIAIS, donné par l'Équation 5.1. Cette fonction, qui est aussi une mesure d'écart cumulé, présentait les mêmes problèmes que le critère de Nash-Sutcliffe.

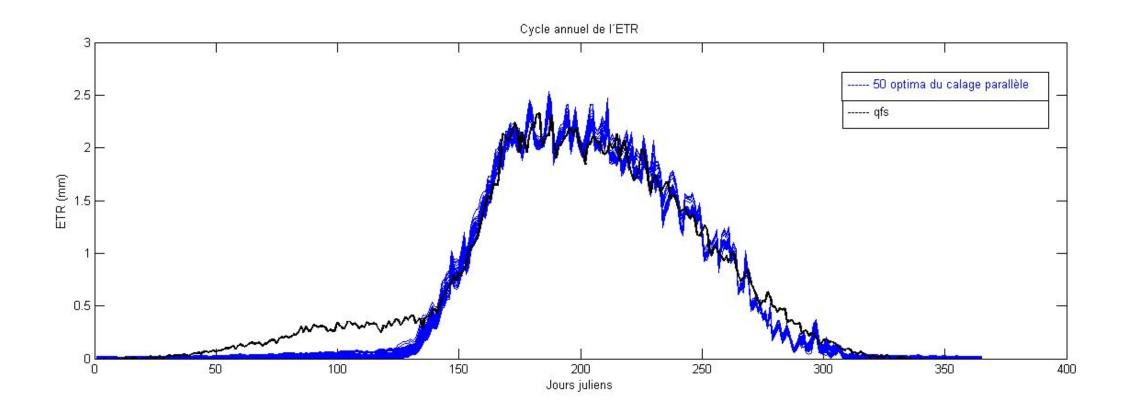


Figure 5.2 Cycle annuel de l'évapotranspiration réelle, ou évapogramme.

$$PBIAS = \left| \frac{\sum_{t=1}^{T} (Q_o^t - Q_s^t)}{\sum_{t=1}^{T} Q_o^t} \times 100 \right|. \tag{5.1}$$

Le deuxième test a donc impliqué une fonction qui, plutôt que de calculer une erreur à chaque pas de temps, consistait à maintenir le volume d'eau évapotranspiré total à une certaine marge du volume de référence. Cette fonction, le biais, se définit comme suit :

$$BIAIS = \left| (\frac{\sum_{t=1}^{T} Q_o^t}{\sum_{t=1}^{T} Q_s^t} - 1) \times 100 \right|. \tag{5.2}$$

De plus, afin d'éviter que le volume plus important en été ne prenne trop d'importance comme dans les stratégies précédentes, nous avons choisi d'appliquer la contrainte non pas sur l'ensemble de la série de données, mais bien par parties. Ainsi, la méthode la plus intuitive consistait à contraindre de façon indépendante l'été et le reste de l'année ; l'été étant délimité par dates standardisées pour un certain nombre de projets menés par l'Institut de recherche d'Hydro-Québec (nous référerons à cette méthode comme la division en deux saisons à partir de ce point). Une séparation plus fine, en trois saisons (soit le printemps, l'été et l'automne combinés et l'hiver), a aussi été mise en place afin de voir si la simulation répondrait de façon significative à ce raffinement.

Finalement, une alternative à la fragmentation par saisons qui est parfois utilisée dans les problèmes impliquant le débit est une approche par intensité de flux. Cette méthode inspirée de Madsen (2000), mais développée par L.-A. Leclaire à l'IREQ calcule le débit moyen de la série, ou, dans le cas qui nous intéresse, l'évapotranspiration moyenne de référence (Qfs_{moy}). Chaque jour julien où l'évapotranspiration simulée sera inférieure à la moitié de cette valeur

($\frac{Qfs_{moy}}{2}$) sera considéré comme faisant partie du régime dit faible. Pour une division en deux régimes, les autres jours feront donc partie du régime élevé. Pour une division en trois régimes, une deuxième valeur sera calculée, soit le carré de l'évapotranspiration moyenne de référence $(Qfs_{moy})^2$. Le régime faible sera le même qu'auparavant, le régime moyen comportera tous les jours dont l'évapotranspiration simulée est comprise entre $\frac{Qfs_{moy}}{2}$ et $(Qfs_{moy})^2$ et le régime élevé incluera les valeurs supérieures à la moyenne au carré. Schématiquement, nous avons donc :

$$\text{Régime faible} \leq \frac{Qfs_{moy}}{2} \leq \text{Régime élevé}, \tag{5.3}$$

$$\text{Régime faible} \leq \frac{Qfs_{moy}}{2} \leq \text{Régime moyen} \leq (Qfs_{moy})^2 \leq \text{Régime élevé}. \tag{5.4}$$

Cette stratégie, qui à première vue semblait prometteuse, a été déjouée par le comportement du modèle hydrologique. En effet, la voie la plus facile pour HSAMI était de faire fondre la neige environ deux semaines plus tôt, puis d'évaporer l'eau, plutôt que d'évaporer l'eau contenue dans la neige, comme le montre le cycle annuel de l'évapotranspiration réelle présenté à la Figure 5.3 où Qfs est représenté en bleu et où la division en deux et trois régimes sont respectivement les courbes verte et rouge.

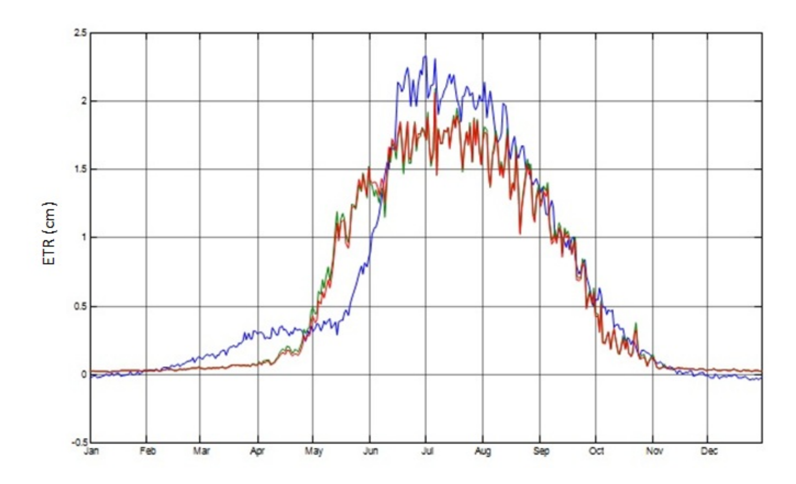


Figure 5.3 Évapogramme des stratégies impliquant le biais pour fonction objectif.

La segmentation de la série semblait tout de même pertinente, mais la fonction n'était toujours pas opportune. Ainsi, l'étape suivante a été de tester la fonction BIAIS_RELATIF (Équation 5.5) avec la même division, en saisons, puis en régimes, malheureusement, avec les mêmes résultats.

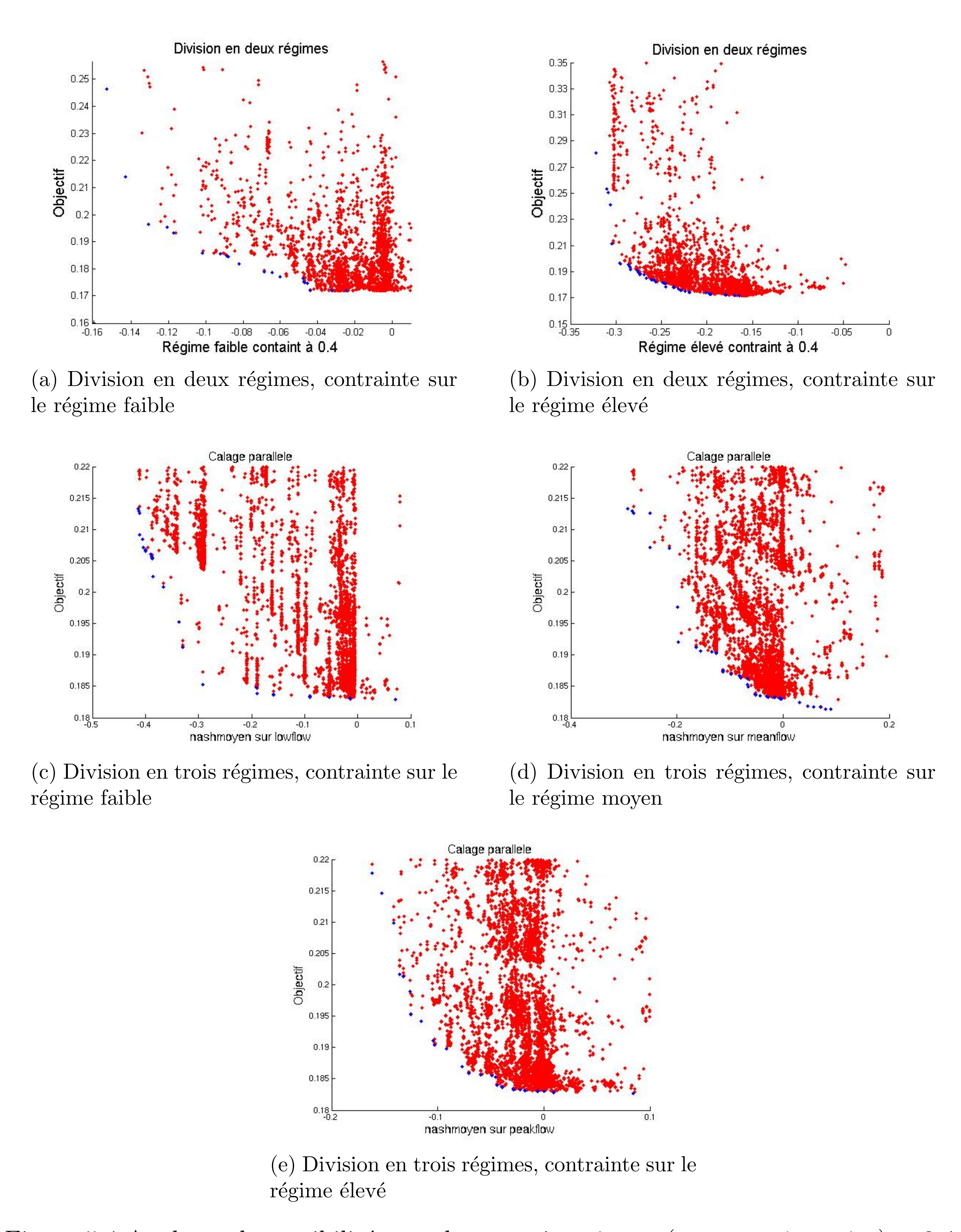
$$BIAIS_RELATIF = \left| \frac{\sum_{t=1}^{T} |Q_o^t - Q_s^t|}{\sum_{t=1}^{T} Q_o^t * nb_total} \times 100 \right| . \quad (5.5)$$

Afin de forcer le modèle à évaporer l'eau au bon moment, le critère de Nash-Sutcliffe appliqué de façon saisonnière (ou par régimes) semblait le plus indiqué, mais l'évapogramme, qui représente la courbe à reproduire, est tellement bruité que l'erreur calculée à chaque pas de temps par cette fonction était énorme. Tout ce chemin nous a donc amené à notre dernière stratégie, soit le critère de Nash-Sutcliffe, appliqué de façon saisonnière ou par régimes, sur l'évapogramme annuel moyen, représentant l'évapotranspiration cumulée par jour julien. La division par saisons a finalement été écartée étant donné la difficulté de transposer une telle stratégie en contexte de changements climatiques, où les divisions des saisons ne seront plus au même moment dû aux changements en termes de précipitations, de températures, de périodes de gel/dégel, etc.

De nouvelles analyses de sensibilité (Figure 5.4) ont donc été effectuées sur les stratégies restantes, soit le critère de Nash-Sutcliffe modifié moyen $(1 - ns_{moy})$sur une période divisée en deux et en trois régimes.

Au départ, une valeur de $(1 - ns)$ inférieure ou égale à 0.4 fut exigée comme contrainte, pour les mêmes raisons que précédemment. À la suite des analyses de sensibilité, il a été convenu de resserrer la contrainte sur le régime moyen ainsi que celle sur le faible régime, pour les deux types de division, à 0.3 (ou 0.7 pour le critère de Nash-Sutcliffe) et celle sur le régime élevé à 0.15 et 0.3 pour les 2 et 3 régimes respectivement (ou 0.85 et 0.7). Une dégradation peu importante de l'objectif a été constatée (augmentation moyenne de 7 % pour la division en deux régimes et de 3 % pour la division en trois régimes), mais notre postulat initial stipulait qu'une certaine détérioration de l'objectif serait tolérée puiqu'une solution avec une valeur de la fonction objectif optimale qui ne respecte pas les contraintes basées sur la physique des processus est moins intéressante qu'une solution moins bonne au niveau du critère de Nash-Sutcliffe mais qui respecte les contraintes physiques. Ainsi, dans l'optique où une stratégie proposant des contraintes plus sévères venait à produire un cycle annuel d'évapotranspiration réelle plus réaliste, cette approche serait retenue comme la nouvelle stratégie de calage en parallèle.

La Figure 5.5 ci-dessous montre donc l'évapogramme annuel des quatre stratégies, soit le critère de Nash-Sutcliffe modifié moyen divisé en deux et en trois régimes, avant et après les analyses de sensibilité effectuées afin de resserrer les contraintes, ainsi que la courbe re-



(a) Division en deux régimes, contrainte sur le régime faible

(b) Division en deux régimes, contrainte sur le régime élevé

(c) Division en trois régimes, contrainte sur le régime faible

(d) Division en trois régimes, contrainte sur le régime moyen

(e) Division en trois régimes, contrainte sur le régime élevé

Figure 5.4 Analyses de sensibilité pour la contrainte $1 - ns(moyen\ saisonnier) \leq 0.4$.

présentant l'évapotranspiration simulée par le modèle climatique, à titre indicatif seulement, puisqu'elle n'a pas été simulée à partir des mêmes données météorologiques. Le but de cette démarche était de reproduire l'allure générale de la courbe afin d'encadrer le processus d'éva-

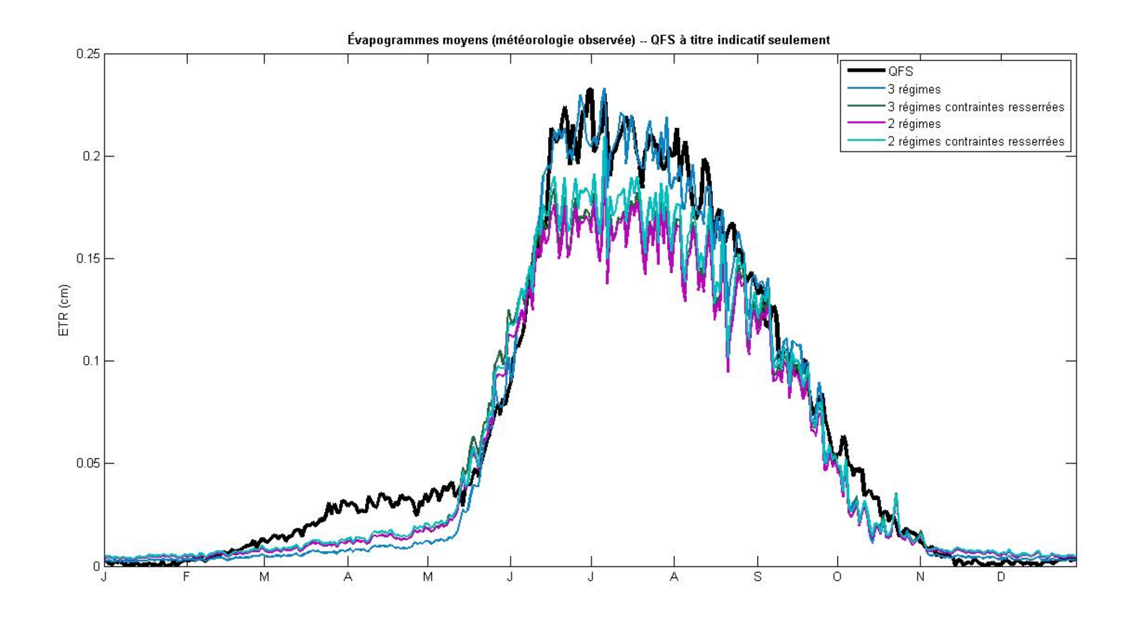


Figure 5.5 Évapogramme des stratégies impliquant deux et trois régimes.

potranspiration, et non pas de calquer les valeurs journalières du modèle climatique. De cette figure, il a été possible d'extraire la stratégie de calage parallèle retenue, soit le critère de Nash-Sutcliffe appliqué sur l'évapogramme annuel cumulé et divisé en deux régimes d'évapotranspiration, soit le régime faible, inférieur à la moitié de l'évapotranspiration moyenne simulée par le modèle climatique ($\frac{Qfs_{moy}}{2}$) et le régime élevé, supérieur à cette même valeur. Les stratégies formées de trois régimes n'avaient pas pour effet de faire augmenter l'évapotranspiration entre mars et la mi-mai. De plus, le décalage du régime élevé, observé entre la mi-mai et la mi-juin, était plus prononcé que pour les stratégies impliquant deux régimes. Finalement, pour les deux stratégies impliquant deux régimes, la sous-estimation de la pointe était moins importante pour la stratégie avec contraintes resserrées, et c'est donc les valeurs de contraintes associées qui ont été retenues, soit 0.3 et 0.15 pour les régimes faible et élevé respectivement, en $1 - ns$.

5.1.2 Analyses de la VNS

Suite à des tests préliminaires réalisés précédemment par Louis-Alexandre Leclaire et moi-même (Leclaire (2011)), il ressortait que l'utilisation de l'option VNS_SEARCH semblait améliorer de beaucoup la performance de NOMAD sur le présent problème. La VNS est une méthode de recherche dans l'espace reposant sur la métaheuristique Variable Neighborhood

Search (détaillée à la Section 2.1.2) et est particulièrement utile lorsque le problème à optimiser comporte un grand nombre d'optima locaux ou un domaine fractionné, comme c'est le cas ici (voir Duan *et al.* (1992) pour les caractéristiques du problème d'optimisation d'un modèle hydrologique). De plus, puisque l'ajout de contraintes permet de réduire l'espace de recherche, il semblait probable que le degré d'utilisation de cet algorithme (en pourcentage du nombre d'évaluations totales) diminuerait d'une stratégie de calage à l'autre. Afin de confirmer cette hypothèse et de choisir le degré d'utilisation de la VNS le plus performant pour chaque stratégie de calage, des tests plus rigoureux furent réalisés. Ainsi, chacune des trois stratégies de calage (traditionnel sans contrainte, traditionnel avec contrainte et parallèle avec contrainte) fut exécutée cinquante fois (afin de diminuer l'impact de la variabilité des résultats induite par le choix des options de NOMAD ainsi que par la forme du problème, tel que décrit dans la Section 4.2.1) pour six degrés d'utilisation de la VNS, soit 0 % (ou VNS désactivé), 25 %, 50 %, 75 %, 90 % et 95 % ; ces stratégies pouvant être raffinées si l'analyse des résultats en démontrait le besoin. L'ensemble des résultats fût regroupé en trois profils de performance, soit un par stratégie de calage, présentés aux Figures 5.6, 5.7 et 5.8.

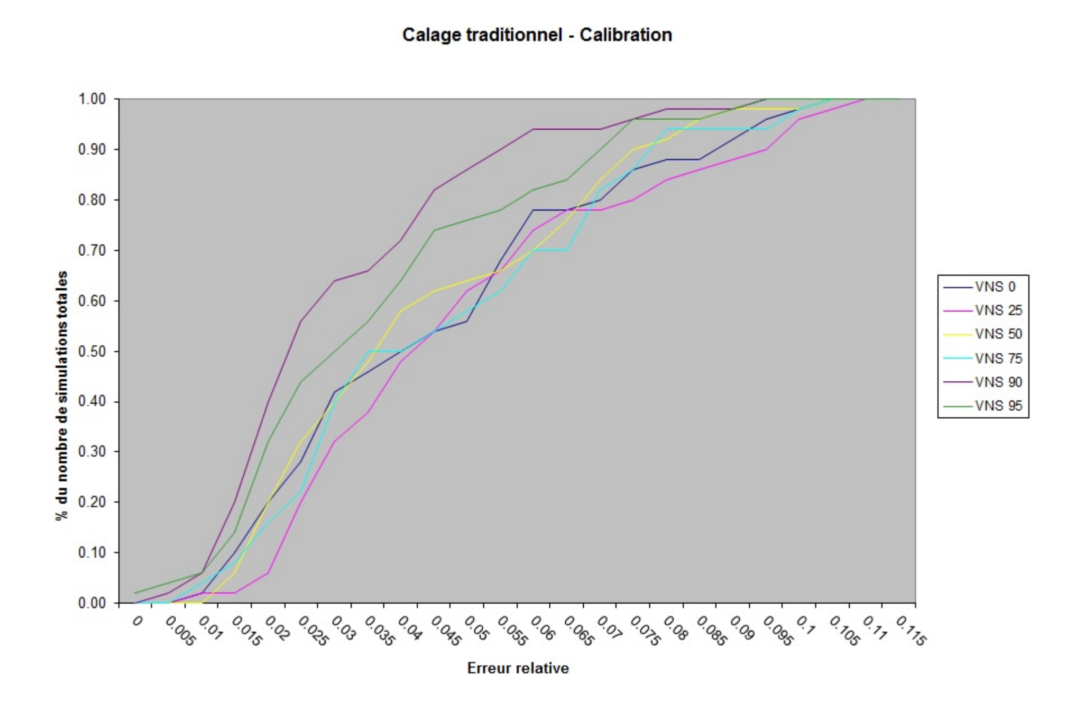


Figure 5.6 Performance des stratégies d'utilisation de la VNS avec un calage traditionnel

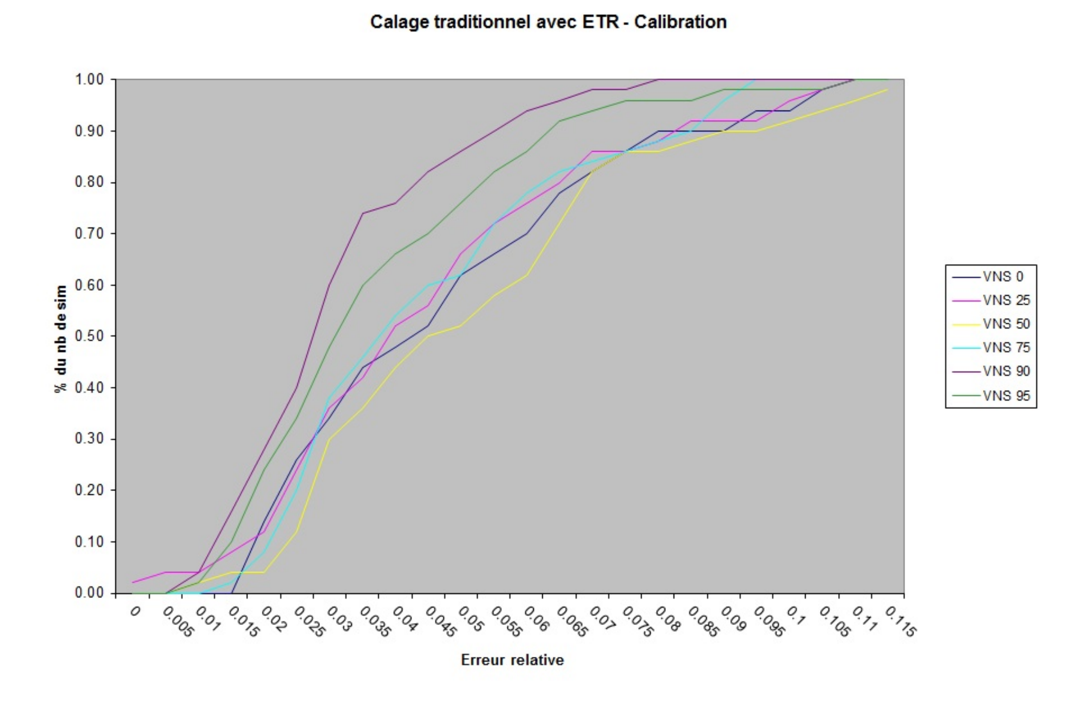


Figure 5.7 Performance des stratégies d'utilisation de la VNS avec un calage traditionnel avec bilan sur l'ETR

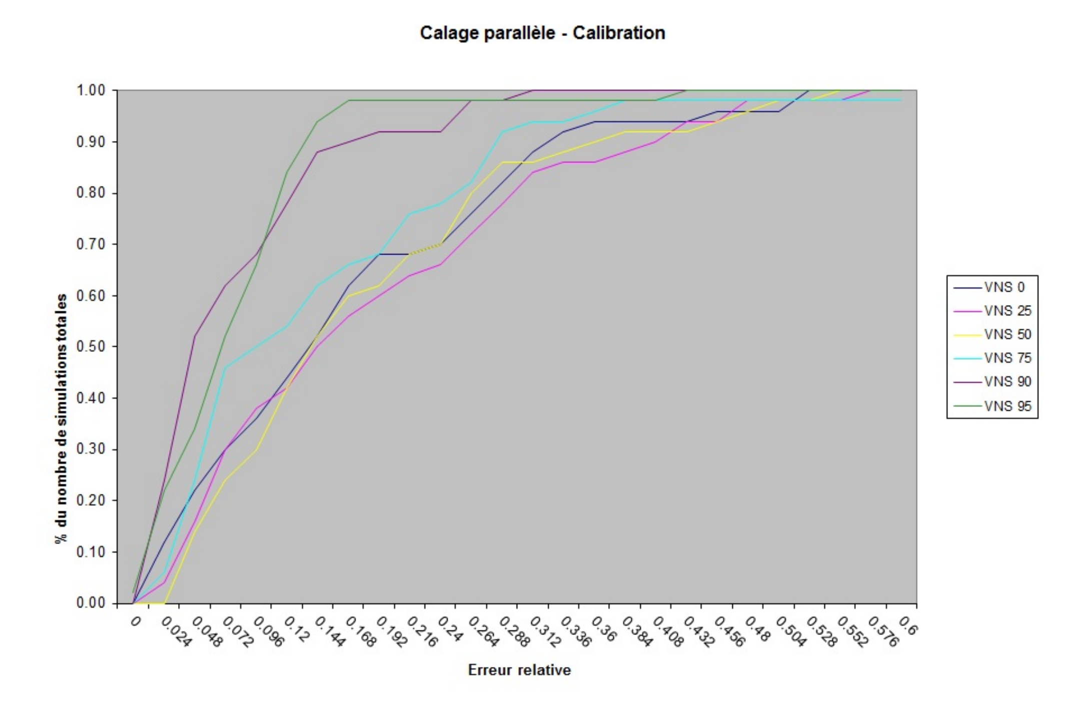


Figure 5.8 Performance des stratégies d'utilisation de la VNS avec un calage parallèle

Introduction aux profils de performance

Les profils de performance ont été conçus par Dolan et Moré (2002) afin de pallier au manque d'outils disponibles afin de comparer efficacement plusieurs algorithmes entre eux. Une technique couramment utilisée consistait à traiter la performance moyenne ou cumulative de chaque algorithme. Le principal inconvénient d'une telle méthode est qu'un petit nombre de problèmes très difficiles influençait grandement les résultats obtenus. De plus, les problèmes n'ayant pu être résolus dans un délai raisonnable devaient être exclus de la comparaison, avantageant les algorithmes les plus robustes, ou encore une pénalité arbitraire leur était attribuée. Plusieurs autres méthodes ont vu le jour, notamment un classement des algorithmes en trois rangs ou par quartiles, mais l'information quant à la progression de la performance desdits solveurs était ainsi perdue. Une approche par ratio par rapport à la meilleure résolution introduite par Billups *et al.* (1997) a conduit à la présente méthode, définie par une fonction de distribution cumulative pour une mesure de performance choisie.

Les profils de performance servent donc à comparer entre elles un certain ensemble S de stratégies différentes sur I problèmes indépendants. Cet outil est donc assez flexible et peut être mis à profit dans bon nombre de situations. Dans le cas présent, ce sont six stratégies d'utilisation de la VNS qui seront comparées entre elles sur leur performance dans la résolution de cinquante instances du problème de calage. En somme, pour chacune des trois stratégies de calage (traditionnel, traditionnel avec bilan ETR et parallèle), ce sont 300 simulations qui seront confrontées à l'aide d'une variable $t_{i,s}$ représentant la valeur de l'objectif obtenue pour le problème i sous la stratégie s. Pour chaque stratégie de calage, le plus petit $t_{i,s}$, ou la meilleure solution, trouvé, toutes stratégies de VNS confondues, sera considéré comme la référence et un ratio de performance, $r_{i,s}$ viendra mesurer l'écart à cette valeur, ou l'erreur relative, de la façon suivante :

$$r_{i,s} = \frac{t_{i,s} - \min\{t_{i,s} : s \in S\}}{\min\{t_{i,s} : s \in S\}}. \tag{5.6}$$

Le graphique est conçu de façon à présenter le nombre de problèmes résolus, pour chaque stratégie $s \in S$, à un certain écart de la meilleure solution, quantitifié par l'erreur relative. Ainsi, la courbe la plus près du coin supérieur gauche du graphique représente la meilleure stratégie, ou celle ayant résolu la plus grande part des problèmes ayant les meilleures solutions. Afin d'illustrer le concept, un exemple comportant deux stratégies, a et b, chacune ayant résolu cinq problèmes, sera présenté ici. Les résultats obtenus apparaissent au Tableau 5.1.

En calculant les erreurs relatives à l'aide de l'Équation 5.6, on obtient le Tableau 5.2. Par

Tableau 5.1 Données de l'exemple académique.

	Stratégie a	Stratégie b
Test 1	5	9
Test 2	8	5
Test 3	4	4
Test 4	1	6
Test 5	6	6

exemple, pour la première valeur du tableau, l'erreur relative se calcule de la façon suivante :

$$r_{i,s} = \frac{5-1}{1} = 4. \quad (5.7)$$

L'étape subséquente consiste à diviser l'étendue totale en un certain nombre d'intervalle, en se servant des valeurs minimale et maximale d'erreur relative. Pour l'exemple présenté ici, les valeurs allant de 0 à 8, des pas d'une unité ont été choisis. Il est alors possible de déterminer, pour chaque pas, combien de problèmes ont été résolus par chaque stratégie (Tableau 5.3). Finalement, il suffit de diviser le nombre de problèmes résolus à chaque intervalle par le nombre total de problèmes (Tableau 5.4) et de tracer le graphique représentant le pourcentage de problèmes résolus à chaque niveau d'erreur relative. Le profil de performance lié au problème académique est présenté à la Figure 5.9. Il permet de conclure à la supériorité de la stratégie a, puisque celle-ci a non seulement trouvé la meilleure solution, mais a aussi trouvé un plus grand nombre de solutions ayant un faible écart avec la meilleure d'entre elles que la stratégie b.

Les profils de performance sont couramment utilisés en recherche opérationnelle (par exemple par Moré et Wild (2009)), mais le sont beaucoup moins en hydrologie, où les diagrammes en boîtes (*boxplot*) sont généralement préférées. Celles-ci seront donc tout de même présentées en Annexe F pour le lecteur intéressé, mais les grandes tendances sont parfois plus difficiles à identifier.

Tableau 5.2 Erreurs relatives de l'exemple académique.

	Stratégie a	Stratégie b
Test 1	4	8
Test 2	7	4
Test 3	3	3
Test 4	0	5
Test 5	5	5

Tableau 5.3 Nombre de problèmes résolus pour chaque intervalle d'erreur relative.

Erreur relative	Stratégie a	Stratégie b
0	1	0
1	1	0
2	1	0
3	2	1
4	3	2
5	4	4
6	4	4
7	5	4
8	5	5

Tableau 5.4 Pourcentage de problèmes résolus par rapport au nombre total pour chaque intervalle d'erreur relative.

Erreur relative	Stratégie a	Stratégie b
0	0.2	0
1	0.2	0
2	0.2	0
3	0.4	0.2
4	0.6	0.4
5	0.8	0.8
6	0.8	0.8
7	1	0.8
8	1	1

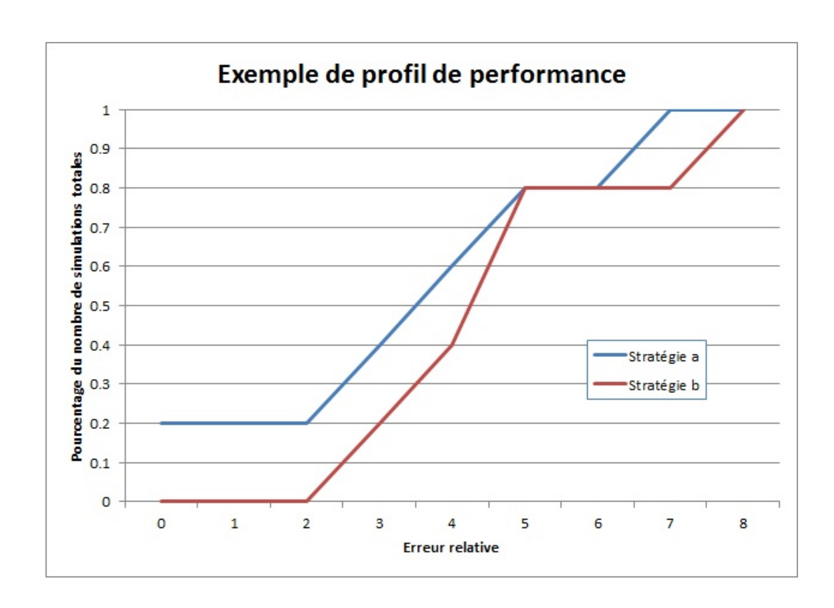


Figure 5.9 Exemple de profil de performance.

5.1.3 Analyse des profils de performance

En se référant à la Figure 5.6, il ressort que la meilleure solution des 300 simulations du calage traditionnel a été trouvée par la stratégie employant un degré de VNS de 95 %. Par contre, globalement, l'emploi de la VNS à 90 % a clairement montré de meilleurs résultats. Par exemple, cette stratégie a résolu 56 % des problèmes (ou 28 des 50) dans un écart de 2.5 % de la meilleure solution, alors que la VNS utilisée à 95 % en a résolu 44 % (ou 22). Le lecteur remarquera que la performance de la stratégie employant la VNS à 90 % est constamment supérieure aux autres, mis à part la section inférieure à 1 % dû au fait que la meilleure solution a été trouvée par une autre stratégie. Celle-ci arrive d'ailleurs bonne deuxième, alors que les autres stratégies ne montrent pas de tendance claire entre elles, mais sont clairement moins performantes que les deux citées plus haut. L'analyse de la deuxième figure (Figure 5.7) impliquant le calage traditionnel avec bilan sur l'ETR est en tout point identique à celle-ci : la meilleure solution a été obtenue par la stratégie utilisant un degré de VNS de 95 % qui se classe deuxième, derrière la stratégie utilisant la VNS à 90 %. Les autres stratégies ne sont pas aussi performantes et peuvent difficilement être différenciées. Finalement, la tendance de la Figure 5.8 est un peu moins évidente, puisque les courbes représentant la VNS utilisée à 90 et à 95 % se croisent. Il est possible de conclure de cette observation que les performances des deux stratégies sont assez proches l'une de l'autre. La stratégie employant la VNS à 90 % a toutefois été préférée puisque, en deça de 10 % d'erreur relative, celle-ci domine la seconde stratégie, ce qui garantit de conserver les meilleures d'entre les meilleures solutions.

Les profils de performance sont donc très clairs sur le choix d'un degré d'utilisation de la VNS de 90 % pour chacune des stratégies de calage. Les résultats obtenus par ces analyses viennent infirmer nos hypothèses selon lesquels le degré de VNS devrait diminuer avec l'ajout de contraintes. L'explication la plus plausible est que le domaine auparavant très vaste a effectivement été réduit par l'ajout de contraintes. Par contre, plutôt que de servir l'optimiseur, le domaine résultant est ressorti fragmenté en plusieurs sections non connexes. Ainsi, un optimiseur tel que NOMAD, opérant par recherche directe sur une grille, peut se retrouver coincé dans une région isolée et donc nécessiter un niveau plus élevé d'utilisation de la VNS afin d'opérer plus librement d'une section à l'autre.

5.2 Comparaison des stratégies de calage

Afin de parvenir à une conclusion quant au choix d'une stratégie de calage pouvant être employée en contexte de changements climatiques, les résultats obtenus par cinquante optimisations pour chacune des trois stratégies de calage [1] seront confrontés au niveau de la variabilité de la paramétrisation, de la robustesse des solutions ainsi que du point de vue hydrologique, notamment dans la représentation de l'évapotranspiration et des débits.

5.2.1 Analyse de la variabilité des paramètres

Lors de tests préparatoires à cette recherche, une forte variabilité des paramètres fut observée en calage traditionnel. Au fil des optimisations, certains paramètres se déplaçaient d'une extrémité à l'autre de la gamme des valeurs comprises entre les bornes. Puisque l'utilisation opérationnelle du modèle, pour les projections en changements climatiques mais aussi pour les prévisions d'apports à court et moyen terme, exige une paramétrisation unique (Efstratiadis et Koutsoyiannis (2010)), une telle variabilité induit une instabilité des résultats peu souhaitable. Il serait donc espéré que les stratégies de calage proposées réduisent les zones paramétriques prometteuses afin d'augmenter la confiance accordée au choix d'un ensemble de paramètres.

La valeur obtenue pour chacun des paramètres, pour les cinquante optimisations et pour chacune des stratégies de calage, est présentée à la Figure 5.10 à titre informatif. Elle ne sera pas détaillée puisque le nombre d'informations qu'elle contient est colossal et dépasse le cadre de ce mémoire. Toutefois, il est intéressant de voir l'impact de la stratégie de calage parallèle sur les paramètres 1 et 2, régissant l'évapotranspiration. Le paramètre 1 permet d'estimer l'évapotranspiration potentielle de la période estivale tandis que le paramètre 2 multiplie le paramètre 1 afin d'estimer l'évapotranspiration potentielle en période hivernale. L'effet du calage parallèle sur le paramètre 2 est fulgurant : en contraignant le modèle afin de mieux représenter l'évapotranspiration en période hivernale, il est forcé à aller dans une zone paramétrique qui était auparavant inexplorée. Les conséquences sur le paramètre 1 ne sont pas aussi notoires, mais il reste que la boîte représentant le calage parallèle est plus basse et moins étendue que celles représentant les autres stratégies. Ce paramètre est forcé au centre de l'intervalle délimité par ses bornes, car, étant impliqué dans le calcul de l'évapotranspiration en périodes estivale et hivernale, il est tenaillé entre les deux contraintes.

1. Pour cette section et les suivantes, les stratégies de calage seront parfois désignées par des numéros afin de ne pas alourdir les figures. Ainsi, 1 correspondra au calage traditionnel, 2 au calage traditionnel avec bilan sur l'ETR et 3 au calage parallèle.

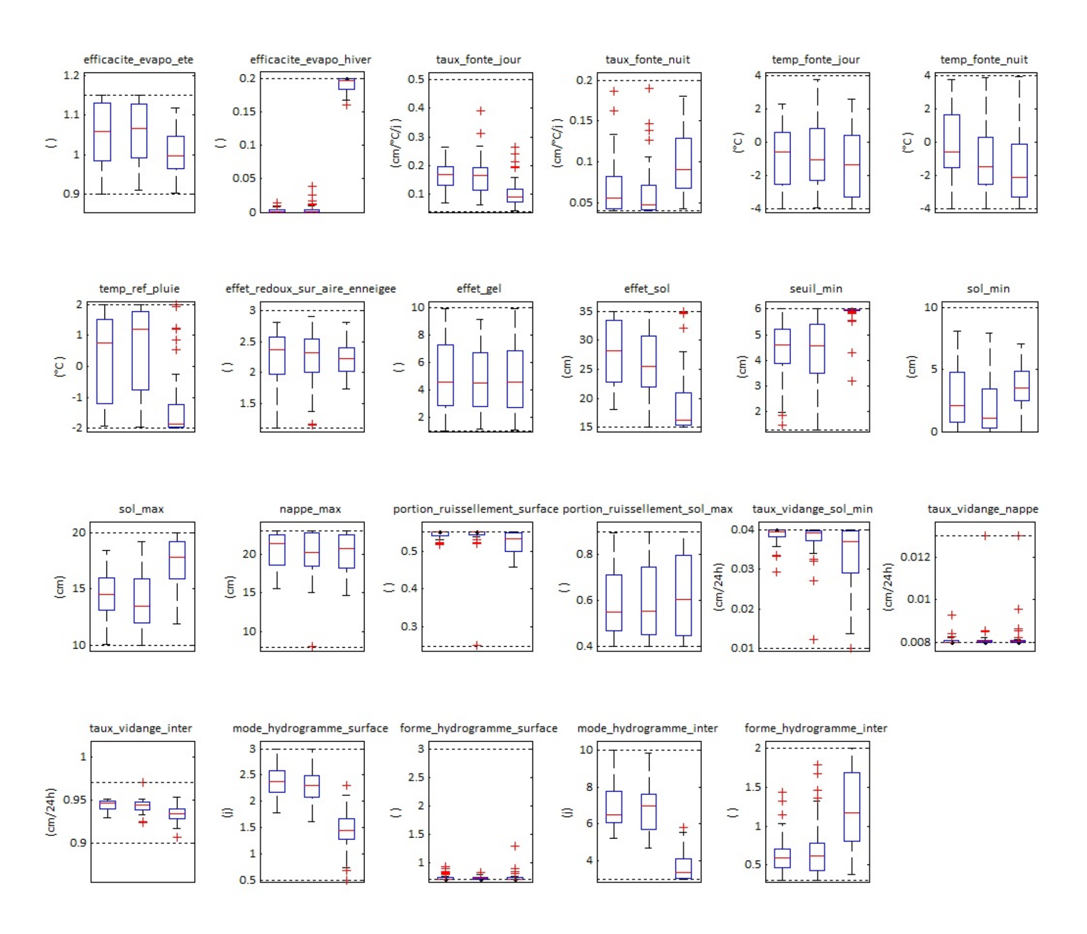


Figure 5.10 Diagrammes en boîtes des paramètres pour chacune des trois stratégies de calage.

Puisque l'écart entre les bornes n'est pas le même pour tous les paramètres, une mise à l'échelle a dû être effectuée avant de pouvoir calculer une variabilité globale pour chaque stratégie. La formule suivante a donc été appliquée sur chaque paramètre des cinquante optima des trois stratégies de calage :

$$ech = \frac{(par - inf_{par})}{(sup_{par} - inf_{par})}. \tag{5.8}$$

où ech représente le paramètre mis à l'échelle, par correspond à la valeur du paramètre de départ, inf_{par} et sup_{par} sont les bornes inférieure et supérieure du paramètre étudié.

Il n'était pas pertinent de regarder la mobilité individuelle de chaque paramètre étant donné la forte interdépendance de certains d'entre eux. L'augmentation de la variance d'un paramètre pourrait ainsi cacher la diminution de celle d'un autre. De plus, il est important de noter que certains paramètres ne sont activés que dans certaines situations extrêmes qui n'étaient pas présentes dans les données utilisées pour le calage. Par conséquent, la fonction objectif était indifférente à la valeur de ces paramètres et leur variabilité est souvent élevée. Cela est cependant le cas pour toutes les stratégies de calage.

Ayant maintenant des bornes communes (0 et 1), les paramètres et leur variabilité peuvent être confrontés. La variance a donc été calculée pour chaque paramètre, et a ensuite été sommée, pour une comparaison globale des stratégies. Les valeurs obtenues sont de 0.9655 pour le calage traditionnel, 1.1137 pour le calage traditionnel avec bilan sur l'ETR et 0.9791 pour le calage parallèle. Le calage traditionnel avec ETR est donc loin derrière les autres stratégies. Surprenamment, le calage parallèle présente une instabilité globale des paramètres légèrement plus importante que celle du calage traditionnel. À titre indicatif seulement, en regardant chaque paramètre individuellement, on peut voir que la variabilité de quinze des vingt-trois paramètres diminue avec l'utilisation du calage parallèle (les variances individuelles et globales sont présentées au Tableau 5.5). De plus, des huit paramètres restant, deux n'influent pas sur la valeur de la fonction objectif (Said (2006)). Il est impossible d'affirmer que les paramètres sont globalement plus orientés sous le calage parallèle à l'aide de ces informations, pour les raisons mentionnées plus haut. Il serait toutefois intéressant, dans de futurs travaux, d'investiguer les six paramètres restants au niveau de leur dépendance à d'autres paramètres, mais aussi de voir comment orienter les processus hydrologiques associés à ces paramètres. En effet, HSAMI pourrait tenter de forcer un processus afin de balancer le niveau d'évapotranspiration demandé. En ciblant et en contraignant ce phénomène, les paramètres pourraient ressortir globalement plus orientés.

5.2.2 Analyse de la robustesse

La robustesse d'une paramétrisation est définie comme étant sa capacité à bien modéliser l'ensemble du régime hydrologique produit à l'aide d'un ensemble de données météorologiques n'ayant pas servi à sa calibration. Il peut s'agir de la moitié des données historiques utilisées pour la validation (avec les débits observés comme référence), mais aussi de données futures (où la robustesse du modèle hydrologique sera estimée selon sa capacité à reproduire les variables de processus hydrologiques d'un modèle climatique). À performance égale en calibration, une paramétrisation robuste sera préférée puisque la modélisation des apports en

Tableau 5.5 Variance des optima.

Paramètre	Traditionnel	Trad. ETR	Parallèle
Paramètre 1	0.10219023	0.086450685	0.041514643
Paramètre 2	0.000210194	0.001190152	0.002480106
Paramètre 3	0.011934586	0.018483887	0.010520919
Paramètre 4	0.047829397	0.038742256	0.056366452
Paramètre 5	0.052401831	0.062063055	0.062112813
Paramètre 6	0.08258015	0.084492748	0.084374989
Paramètre 7	0.119558113	0.117797582	0.072870954
Paramètre 8	0.058113835	0.051500004	0.016721082
Paramètre 9	0.08322342	0.072161954	0.077492192
Paramètre 10	0.074305293	0.08392272	0.087666652
Paramètre 11	0.061104591	0.078447641	0.009252616
Paramètre 12	0.057518086	0.045875535	0.027454786
Paramètre 13	0.044734954	0.064168261	0.049015547
Paramètre 14	0.020420433	0.034335791	0.025304384
Paramètre 15	0.000784245	0.019597931	0.007741663
Paramètre 16	0.073672978	0.109143719	0.123346436
Paramètre 17	0.004602986	0.021708814	0.076686284
Paramètre 18	0.001497183	0.019604276	0.021229929
Paramètre 19	0.00617982	0.010082895	0.01656564
Paramètre 20	0.012419599	0.019291444	0.019810553
Paramètre 21	0.000541216	0.000193251	0.001481042
Paramètre 22	0.028960769	0.027538041	0.011294091
Paramètre 23	0.020756646	0.046861972	0.077809575
Somme	0.965540553	1.113654616	0.979113349

mode prévisionnel sera vraisemblablement plus crédible.

Ainsi, en comparant les différentes stratégies au niveau de leurs résultats en calibration et en validation présentés à la Figure 5.11, la première chose qui saute aux yeux est que toutes les méthodes réussissent mieux en validation qu'en calibration. L'explication est plutôt simple : les années ayant servi à la calibration, soit les années paires, sont considérées comme étant plutôt humides, tandis que les années de validation sont plus sèches. En effet, les précipitations moyennes annuelles des années paires et impaires sont de 840 mm et 810 mm respectivement. Plusieurs articles, notamment Gupta *et al.* (1996), Gupta *et al.* (1998) et Gan *et al.* (1997), ont conclu qu'un modèle calibré en régime humide était mieux entraîné qu'un modèle ayant été étalonné en régime sec, car il est plus apte à savoir réagir aux différents événements météorologiques. D'après Cooper *et al.* (2007), un tel ensemble de données utilisé en calibration aurait d'ailleurs pour effet d'engendrer un espace de recherche plus étendu.

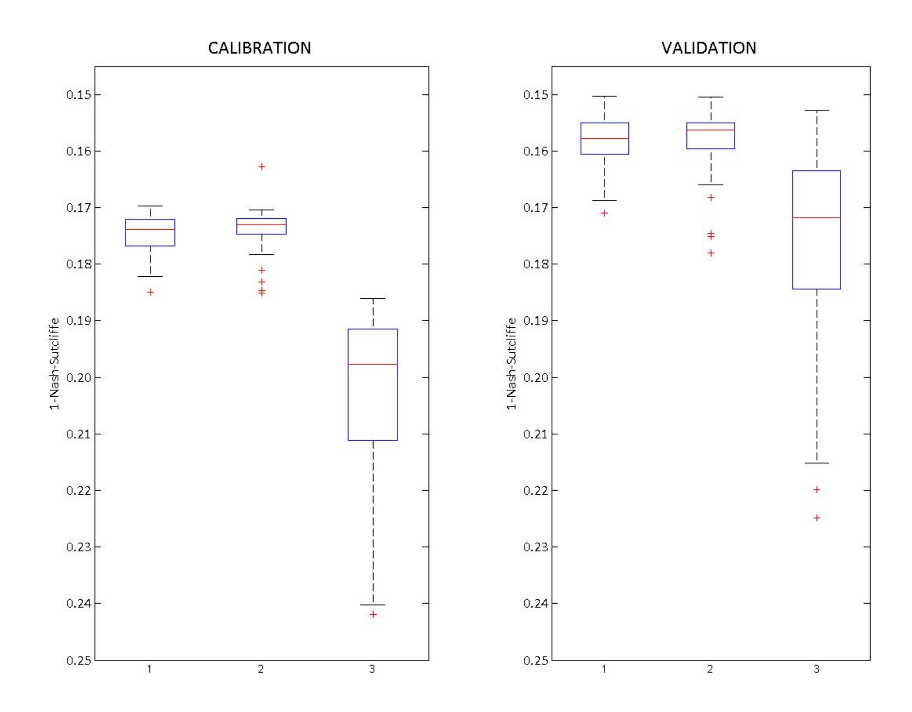
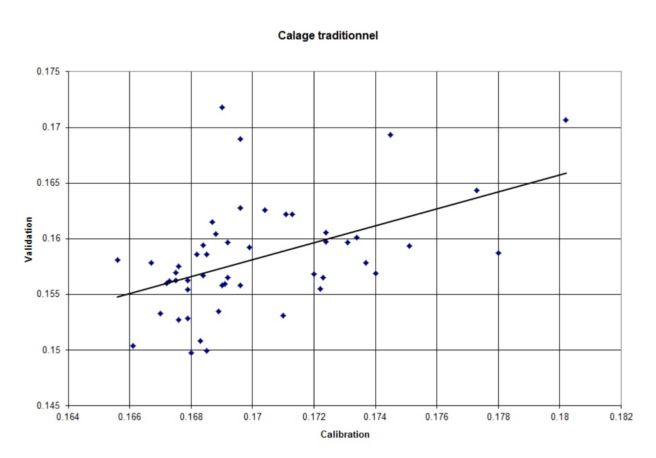
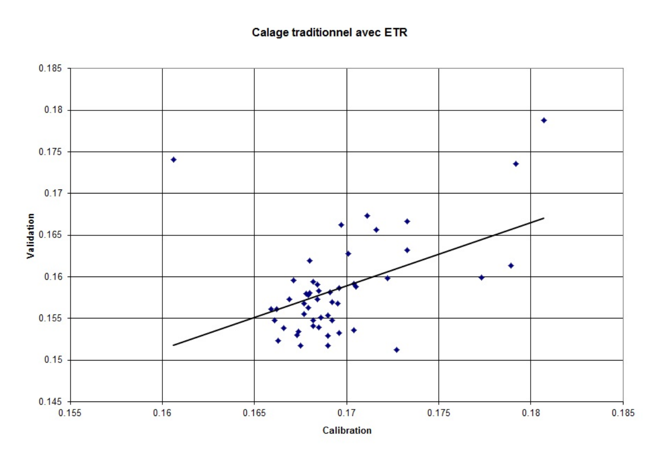


Figure 5.11 Valeur du critère de Nash-Sutcliffe en calibration et en validation pour les trois stratégies de calage.

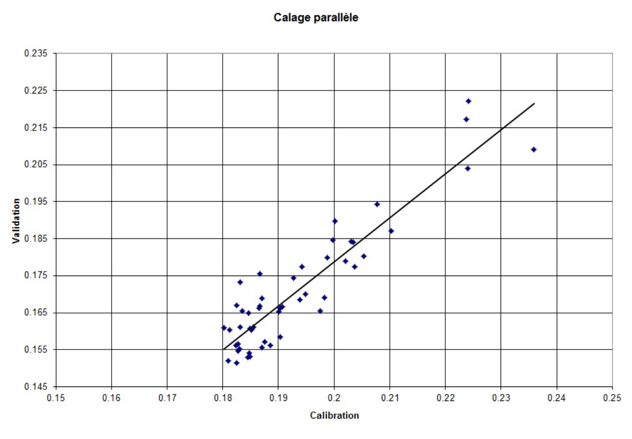
Avant d'aller plus loin dans l'analyse, un questionnement a été soulevé quant aux points extrêmes, ou *outliers*, des diagrammes en boîtes. Il aurait été intéressant de voir s'il existait une certaine corrélation entre les points extrêmes trouvés en calibration et ceux trouvés en validation. Les extremums ont donc été étudiés un à un et une relation a semblé émerger. Afin de clarifier la situation, tous les points, extrêmes ou non, ont été tracés dans un graphique (Figure 5.12) présentant la valeur obtenue en calibration en fonction de celle obtenue en validation, et ce selon la stratégie de calage utilisée. Des tests de corrélation ont aussi été effectués afin de quantifier l'ampleur de la relation reliant ces deux variables. Pour le calage traditionnel et traditionnel avec bilan sur l'ETR, les valeurs de corrélation obtenues sont de 0.4869 et 0.4589 respectivement, alors que pour le calage parallèle, cette valeur grimpe à 0.9179. Comme il sera démontré à l'Annexe G, cette corrélation est réellement une propriété du calage parallèle. Cette constatation est lourde de conséquences, en ce sens où, lors de l'utilisation de cette stratégie de calage, la qualité des solutions est stable sur des ensembles de données météorologiques différents, signe d'une grande robustesse. Ainsi, en choisissant la meilleure solution obtenue en calibration, l'utilisateur sera certain d'avoir aussi les paramètres correspondant à une des meilleures solutions produites en validation.



(a) Corrélation observée pour le calage traditionnel



(b) Corrélation observée pour le calage traditionnel avec bilan sur l'ETR



(c) Corrélation observée pour le calage parallèle

Figure 5.12 Analyse de la corrélation des valeurs obtenues en calibration et en validation.

En retournant à l'analyse de la Figure 5.11, une autre conclusion à tirer est que le calage

parallèle offre, globalement, de moins bons résultats que les autres stratégies. Cependant, en prenant en compte la particularité du calage parallèle à engendrer une corrélation des solutions en calibration et en validation, seule la meilleure solution obtenue par le calage parallèle a de l'intérêt. Ainsi, en confrontant la paramétrisation la plus robuste de la troisième stratégie aux deux autres, force est de constater que celle-ci, en validation, est équivalente à la grande majorité des autres solutions. Notons aussi que l'ensemble des solutions trouvées par le calage parallèle offre de meilleurs résultats en validation qu'en calibration, ce qui est la définition même de la robustesse.

En ce qui concerne la stabilité des solutions, plusieurs constatations peuvent être énoncées. En premier lieu, la deuxième stratégie, calage traditionnel avec bilan sur l'ETR, est moins dispersée que le calage traditionnel. Cela s'expliquerait par le fait que, les observations de Caniapiscau le situant déjà dans l'intervalle de la contrainte, l'espace de recherche a pu être resserré par rapport au calage traditionnel qui effectue ses recherches dans l'espace paramétrique tout entier. Ensuite, il convient de mentionner la plus grande variabilité des solutions obtenues par calage parallèle. Les contraintes imposées par cette stratégie de calage étant beaucoup plus contraignante, le domaine s'est vu être fractionné en plusieurs sous-espaces. Une telle configuration de l'espace paramétrique est difficile à explorer, comme il fut mentionné dans la Section 5.1.2. Ainsi, plusieurs optima locaux furent trouvés par l'algorithme, ce qui a entraîné une plus grande variabilité dans l'ensemble des solutions.

En somme, en ce qui concerne la variabilité et la robustesse, il est difficile de statuer sur une méthode radicalement meilleure que les autres. Les indicateurs sont plutôt équivalents pour l'ensemble des stratégies et, alors que le calage parallèle semble moins performant que les autres en terme de fonction objectif (conséquence que nous étions prêts à accepter), le lien de corrélation obtenu sous le calage parallèle entre les valeurs obtenues en calibration et en validation représente une belle avancée en terme de robustesse. Les analyses hydrologiques qui suivent permettront de faire la part des choses quand à la performance réelle de chacune des stratégies.

5.2.3 Analyses hydrologiques

Les variables hydrologiques d'intérêt seront donc analysées afin d'évaluer leur comportement sous chacune des stratégies de calage en espérant que les différences ressortent et qu'une conclusion puisse être prise quant à la performance des différentes méthodes. Rappelons encore une fois notre hypothèse de départ selon laquelle une solution avec une valeur de la fonction objectif optimale qui ne respecte pas les contraintes basées sur la physique

des processus est moins intéressante qu'une solution moins bonne au niveau du critère de Nash-Sutcliffe mais qui respecte les contraintes physiques. Après avoir regardé seulement la valeur de l'objectif, nous nous attarderons donc ici au respect de la physique des processus hydrologiques, notamment à la représentation du débit et de l'évapotranspiration. Toujours selon notre hypothèse de recherche, il serait normal de constater une dégradation dans le portrait du débit si cela résulte en une image plus fidèle de l'ETR.

Hydrogrammes

Les hydrogrammes ont donc été générés et sont présentés en trois sous-figures référencées 5.13, soit une par stratégie de calage, car les courbes superposées des trois stratégies étaient confondues et il devenait impossible de distinguer quoi que ce soit. C'est un bon signe, en ce sens où, de façon générale, les débits sont assez bien représentés par toutes les stratégies. Cependant, il faut mentionner un problème considérable, soit le retard de la crue par la stratégie de calage parallèle. Cet événement ayant une grande importance, notamment pour la gestion des barrages, ce décalage doit être pris au sérieux. Après investigation, une hypothèse a été proposée concernant une modélisation erronée de la neige. Puisque le modèle subit un déficit de neige, mais que les contraintes du calage parallèle imposent tout de même un certain niveau d'évapotranspiration, HSAMI est forcé de prendre l'eau quelque part et c'est la crue qui en pâtit. Cette piste sera développée à l'Annexe G.2 et viendra effectivement expliquer cet écart. Il n'est cependant pas possible de statuer sur la représentation de la crue par le calage parallèle puisque le problème de modélisation de la neige fait partie des pistes à explorer dans le futur. Malgré tout, la méthode développée ici a été pensée pour une utilisation en contexte de changements climatiques et, pour évaluer les impacts en termes de volume desdits changements, le synchronisme de la courbe a moins d'importance. Dans cette optique, la représentation des débits par la méthode de calage parallèle est considérée comme étant acceptable. Notons que les stratégies de calage traditionnel et traditionnel avec bilan sur l'ETR ne sont pas affectées par cette complication car, les processus hydrologiques ayant toute la latitude voulue, les erreurs s'annulent entre elles. C'est d'ailleurs ce type de comportement qui inquiète quant à la capacité de ces méthodes à bien modéliser les apports, surtout dans un contexte de changements climatiques où les données historiques servant au calage ne sont pas garantes des conditions météorologiques futures.

L'hydrogramme montre aussi une plus grande variabilité entre les solutions pour la stratégie de calage parallèle. Cependant, puisqu'une seule paramétrisation sera retenue au final, et puisqu'il a été observé dans la Section 5.2.2 qu'une forte corrélation unissait les valeurs

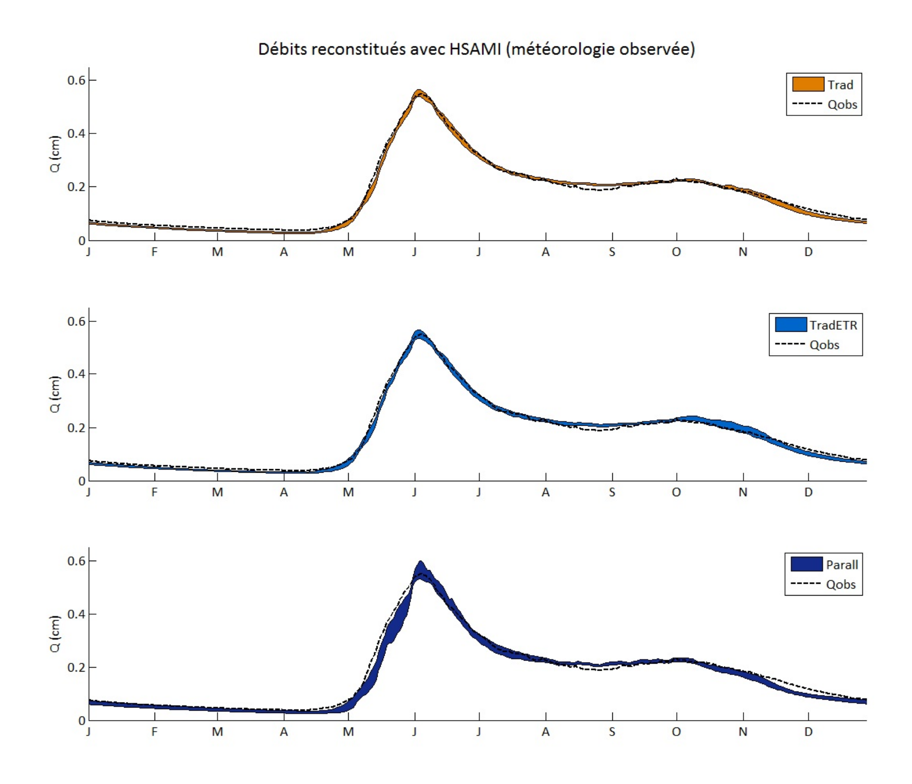


Figure 5.13 Hydrogrammes produits par cinquante simulations pour chaque stratégie de calage avec la météologie observée (krigIREQ).

obtenues en calibration et en validation, nous pouvons supposer que la paramétrisation choisie fournira une des meilleures représentations du débit. La Figure 5.14 montre donc l'hydrogramme associé à la meilleure solution obtenue en calibration pour le calage parallèle confronté au débit observé et prouve que la représentation des apports par la meilleure solution obtenue en calage parallèle est comparable à celle produite par les deux autres stratégies de calage. Notons que le retard de la crue est toujours présent, mais n'est pas plus prononcé que pour l'ensemble des autres solutions obtenues par calage traditionnel et traditionnel avec bilan sur l'ETR.

Évapogrammes

L'impact des différentes stratégies de calage est beaucoup plus marqué sur les évapogrammes. Une première série d'évapogrammes (Figure 5.15) a été générée à partir des données observées. La variable Qfs représentant l'évapotranspiration simulée par le modèle climatique n'est pas présentée, n'étant pas comparable aux cycles d'ETR produits par les trois stratégies de calage. En effet, ceux-ci ayant été simulés par une simulation alimentée par des données

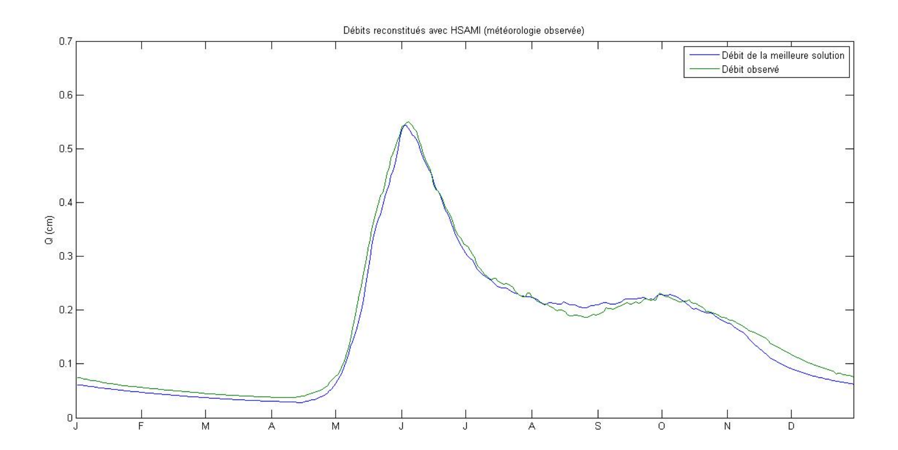


Figure 5.14 Hydrogramme comparant les débits observés à la meilleure solution du calage parallèle.

météorologiques différentes, ce sont deux scénarios possibles du même monde que nous tenterions de comparer.

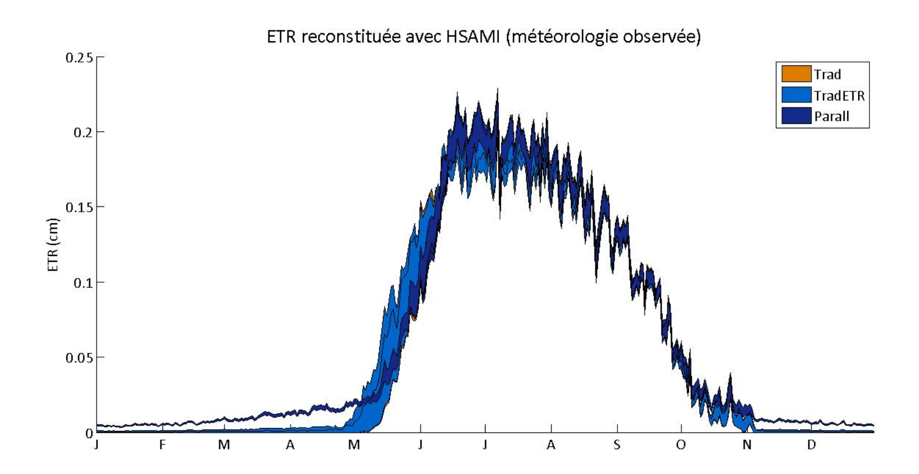


Figure 5.15 Évapogrammes produits par cinquante simulations pour chaque stratégie de calage alimentées par la météorologie observée.

La courbe représentant les cinquante optima du calage traditionnel peut sembler absente, mais elle est presque superposée à celle présentant les optima du calage traditionnel avec bilan sur l'ETR. De cette figure, il est possible de constater que le calage parallèle offre une moins grande variabilité que les deux autres stratégies, et donc que la majorité des solutions obtenues par calage parallèle modélise bien l'évapotranspiration. De plus, en s'attardant aux mois entre novembre et mai, le calage parallèle réussit à augmenter l'évapotranspiration, alors

que le calage traditionnel et le calage traditionnel avec bilan sur l'ETR n'y parviennent pas.

Afin de permettre une comparaison valide avec la variable Qfs, et donc de situer les stratégies quant à leur capacité de modéliser le processus d'évapotranspiration en respectant la physique, HSAMI doit simuler les apports pour le même monde virtuel, et donc avec les mêmes données météorologiques. Il a donc été relancé en mode simulation avec les cent cinquante optima trouvés par les trois stratégies de calage, mais avec les données météorologiques d'*aft* en intrants. C'est donc l'équivalent de la version d'HSAMI dédiée au calcul de la contrainte du calage parallèle qui a été relancé, mais sans optimisation. Les évapogrammes présentés à la Figure 5.17 n'ont pas été superposés afin de faciliter la comparaison avec la courbe représentant la variable Qfs. La superposition des courbes est tout de même présentée à la Figure 5.16 à titre indicatif.

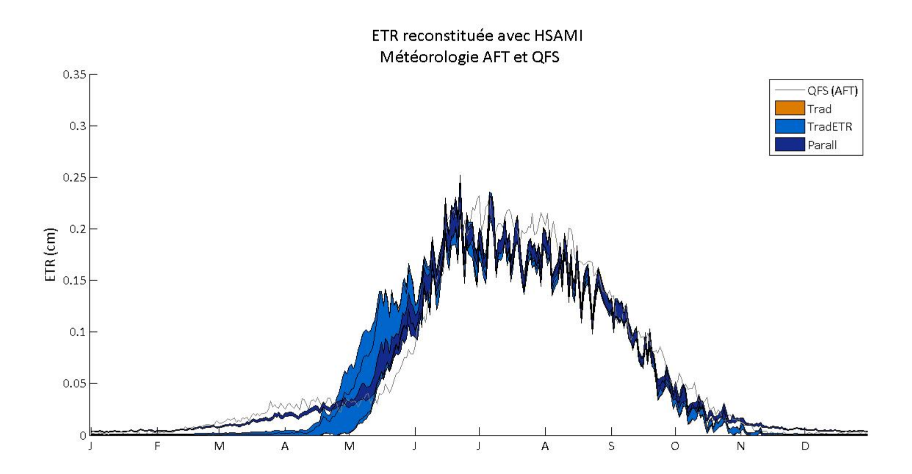


Figure 5.16 Évapogrammes pour les trois stratégies de calage alimentées par les données climatiques d'*aft* (les courbes du calage traditionnel et du calage traditionnel avec ETR sont superposées).

Les conclusions sont les mêmes que pour l'évapogramme précédent, soit que le calage parallèle offre une variabilité moins importante que les autres stratégies et que sa représentation de l'évapotranspiration telle que modélisée par un modèle climatique est beaucoup plus fiable, notamment en ce qui concerne la période hivernale. Pour ce qui est du début de l'été, ou des mois entre mai et juillet, il est aussi possible de remarquer que la courbe liée au calage parallèle est plus près de Qfs, alors que la hausse de l'évapotranspiration simulée par les deux autres stratégies est liée à l'eau accumulée durant les mois précédents.

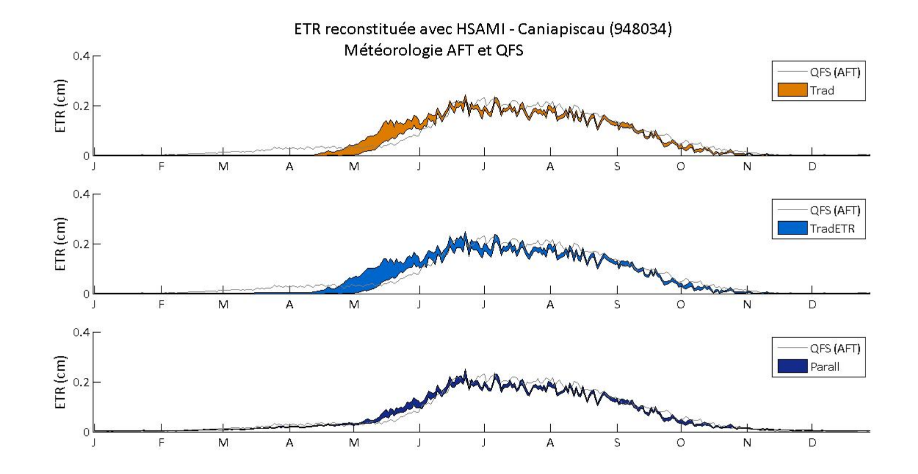


Figure 5.17 Évapogrammes produits par cinquante simulations pour chaque stratégie de calage alimentée par la météorologie d'*aft*.

Respect des contraintes

Toujours afin de comparer les méthodes entre elles, le respect des contraintes imposées à chacune a été vérifié sur les années non utilisées pour l'optimisation, mais aussi sur les deux autres stratégies.

Ratio ETR/P Ainsi, à la Figure 5.18, le ratio $\frac{ETR}{P}$ a été évalué pour chacun des optima. Notons que la ligne pleine représente la valeur calculée à partir des observations uniquement ($1 - \frac{Q}{P} \equiv \frac{ETR}{P}$), que la ligne pointillée indique la valeur de l'Atlas du Canada et que la ligne alternant points et tirets représente la valeur calculée à partir des données météorologiques de la simulation climatique *aft* ($\frac{Qfs}{P}$). Il est donc normal que le ratio de la stratégie de calage traditionnel avec bilan sur l'ETR tende vers la valeur de l'Atlas, alors que la stratégie de calage parallèle devrait chercher à s'élever vers la valeur calculée sur les données d'*aft*. Il est aussi légitime qu'il y ait peu de différences entre les première et deuxième stratégies puisque, pour le bassin Caniapiscau, les observations se situaient déjà dans l'intervalle imposé par la contrainte du calage traditionnel avec bilan sur l'ETR. L'impact entre ces deux stratégies sera plus important pour un bassin versant dont la valeur du ratio $\frac{ETR}{P}$ sera plus éloigné de celle de l'Atlas (voir Annexe G). La paramétrisation du modèle sous la stratégie de calage parallèle permet d'augmenter l'évapotranspiration de façon manifeste (presque 3 % de gain sur le ratio $\frac{ETR}{P}$ par rapport aux autres stratégies). Cependant, il est impossible de statuer sur une stratégie à la vue de cette figure puisque l'on ne sait pas quelle source d'informations, entre l'Atlas du Canada et le modèle climatique, correspond à la réalité.

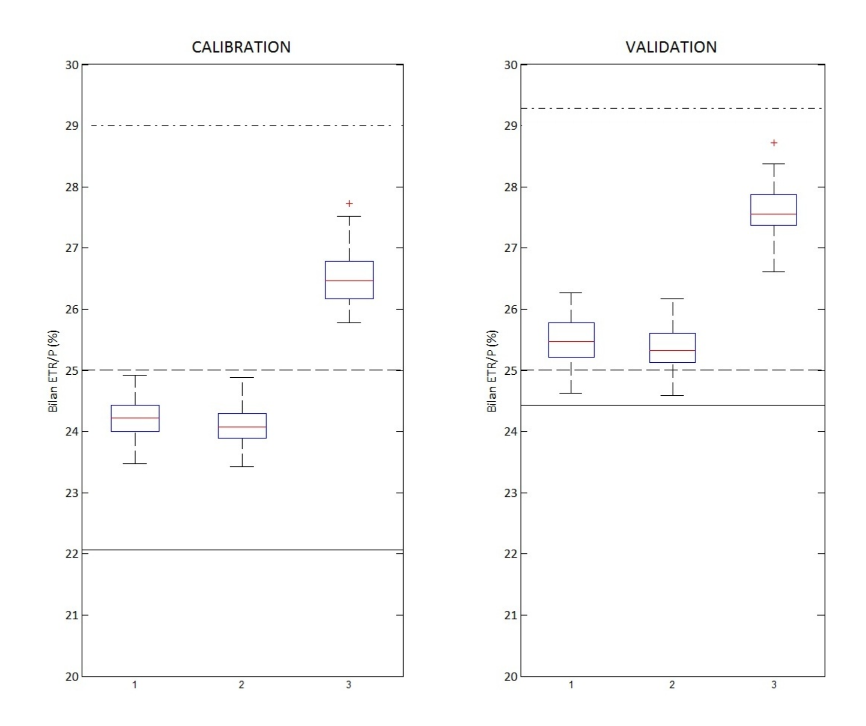


Figure 5.18 Calcul du ratio $\frac{ETR}{P}$ pour les trois stratégies de calage alimentées par les données krigIREQ.

La même expérience a été répétée, mais cette fois en alimentant HSAMI avec les données météorologiques d'*aft*, afin de voir si les stratégies de calage traditionnel et traditionnel avec bilan sur l'ETR sauraient d'adapter. Ces deux stratégies ont été calibrées pour un ratio $\frac{ETR}{P}$ de 25 %, mais, dans ce monde virtuel, le ratio attendu est plus élevé d'un peu moins de 4 % (identifié par le trait noir). La Figure 5.19 montre une légère augmentation du ratio pour les deux premières stratégies, en calibration et en validation. Le calage parallèle est tout de même, et de loin, le plus près de la valeur espérée de presque 29 %. Ainsi, cette stratégie offre non seulement une bonne représentation du cycle de l'ETR, mais elle parvient à une estimation plus réaliste des quantités, contrairement aux deux autres stratégies qui tendent à sous-estimer l'évapotranspiration.

Mesures d'erreur sur les périodes estivale et hivernale Afin de pouvoir évaluer les contraintes liées au calage parallèle, il a été nécessaire de reproduire la méthode utilisée pour tracer les évapogrammes avec Qfs, soit de relancer HSAMI alimenté par les données météorologiques d'*aft* avec les cent cinquante paramétrisations fixées. Avec les séries d'évapo-

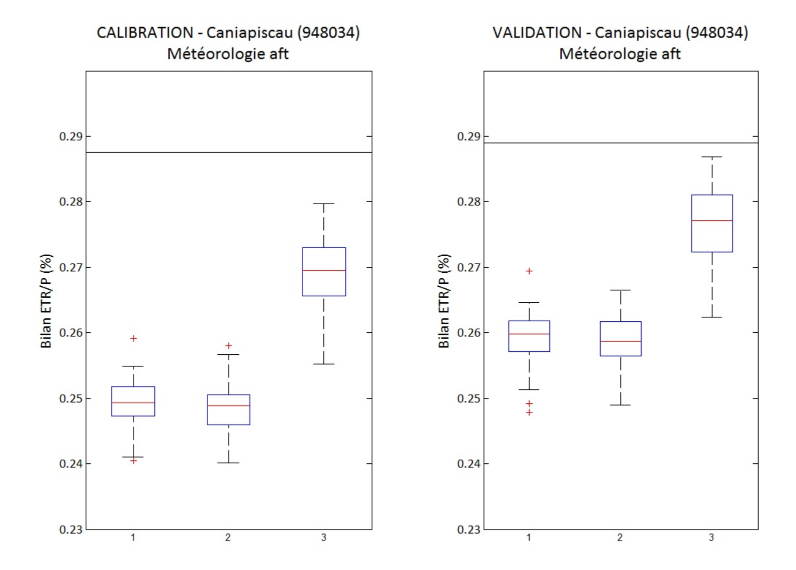


Figure 5.19 Calcul du ratio $\frac{ETR}{P}$ pour les trois stratégies de calage alimentées par les données climatiques, et pour les données météorologiques d'*aft* (trait noir).

transpiration ainsi simulées, il sera possible d'effectuer le calcul du critère de Nash-Sutcliffe [2] avec la variable Qfs, provenant du modèle climatique. Pour la période estivale, présentée à la Figure 5.20, le calage parallèle offre de meilleures solutions que les autres méthodes, et celles-ci sont beaucoup moins variables. Notons cependant que l'évapotranspiration simulée par le calage traditionnel et par le calage traditionnel avec bilan sur l'ETR n'est pas si loin de la réalité telle que représentée par Qfs. Comme il a été mentionné dans la Section 4.1.1, la fonction objectif utilisée pour l'optimisation de ce problème met beaucoup de poids sur les valeurs importantes, et donc sur la crue. Ainsi, même une optimisation sans contrainte se doit de relativement bien représenter cette période, coincidant avec la période estivale en termes d'évapotranspiration. Le modèle offre donc moins de marge de manœuvre durant cette période et le processus est plutôt bien représenté par toutes les stratégies de calage.

La période hivernale, toutefois, n'est pas affectée par cette particularité du modèle. Les résultats de la Figure 5.21 sont donc plus impressionnants. Ainsi, l'ensemble des solutions proposées par les stratégies de calage traditionnel et de calage traditionnel avec bilan sur l'ETR sont négatives, signe qu'une utilisation de la simple moyenne serait plus performante. Le calage parallèle, à l'opposé, arrive à des solutions plus qu'acceptables, avec des valeurs

2. Il est important de noter que c'est le critère de Nash-Sutcliffe, et non $1 - ns$, qui est présenté sur les figures qui suivront. Une valeur près de 1 est donc garante d'une bonne représentation du cycle de l'évapotranspiration, au contraire d'une valeur près de 0. Une valeur négative est signe qu'une utilisation de la simple moyenne serait plus judicieuse que la série ayant été simulée.

de Nash-Sutcliffe de plus de 0.7, et ce autant en calibration qu'en validation. Il est donc possible d'affirmer que la représentation du processus d'évapotranspiration durant la période hivernale par la méthode du calage parallèle est de beaucoup supérieure à celle offerte par les autres méthodes.

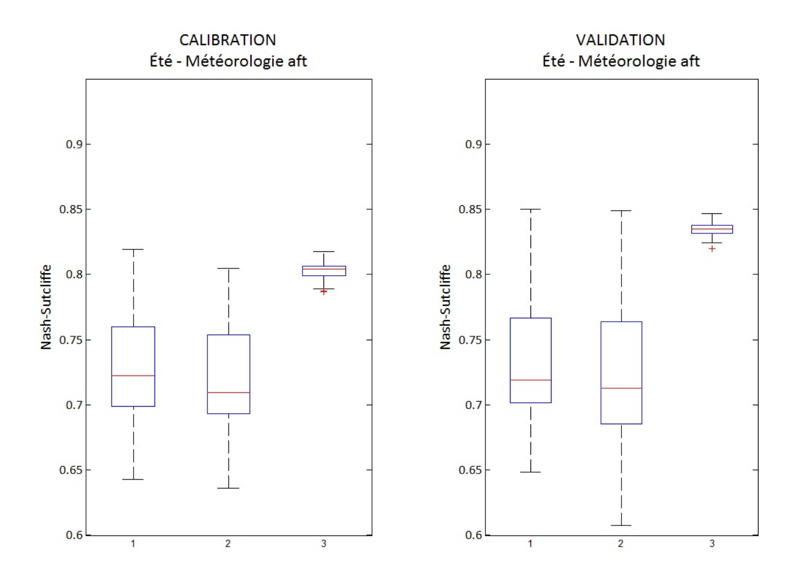


Figure 5.20 Critère de Nash-Sutcliffe sur la période estivale pour les cinquante optimisations de chaque stratégie de calage alimentée par la météorologie d'*aft*.

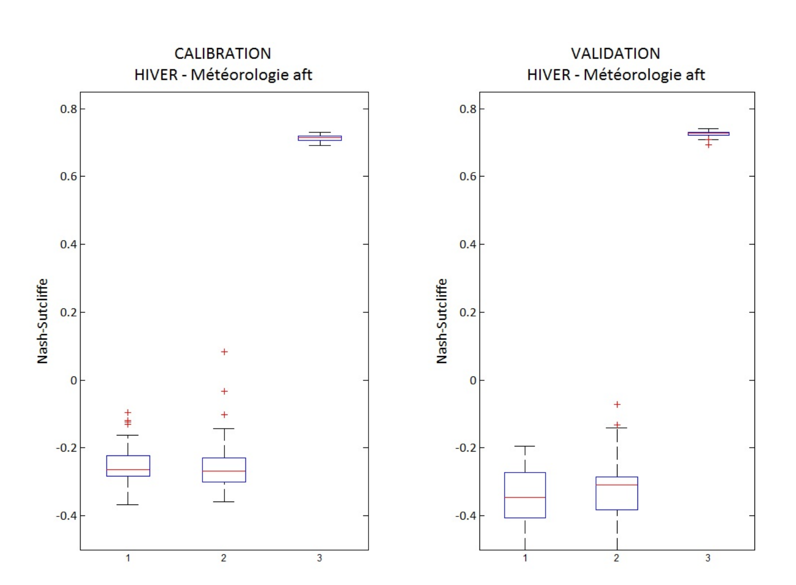


Figure 5.21 Critère de Nash-Sutcliffe sur la période hivernale pour les cinquante optimisations de chaque stratégie de calage alimentée par la météologie d'*aft*.

5.3 Synthèse des résultats

Afin de synthétiser la quantité considérable d'informations contenue dans ce chapitre, les conclusions générales concernant chacune des trois stratégies de calage seront reprises ici. Elles seront ensuite confrontées dans le but de parvenir à une recommandation pour la calibration du modèle HSAMI pour une utilisation en contexte de changements climatiques [3].

5.3.1 Calage traditionnel

Des trois, le calage traditionnel était, avant cette recherche, la seule méthode utilisée chez Hydro-Québec pour l'optimisation des paramètres opérationnels du modèle HSAMI. Malgré des performances satisfaisantes en terme de représentation des débits sur des données historiques, cette méthode soulève des doutes quant à sa capacité de représenter les apports naturels et les processus hydrologiques produits par un contexte hydrologique différent. La grande latitude au niveau de la modélisation des processus hydrologiques permet actuellement de balancer les erreurs entre elles, mais lorsque le modèle ne pourra plus se fier sur les débits passés pour estimer les débits futurs, une modélisation respectant la physique des processus serait souhaitable. Dans les conditions actuelles, il a été démontré que le processus d'évapotranspiration, particulièrement en période hivernale, était excessivement mal modélisé. De fait, l'utilisation de la simple moyenne serait plus judicieuse. Le calage traditionnel n'offre donc aucun avantage par rapport aux autres méthodes pour une calibration en contexte de changements climatiques.

5.3.2 Calage traditionnel avec bilan sur l'ETR

Le calage traditionnel avec bilan sur l'ETR, à tout le moins pour le bassin versant Caniapiscau, offre une performance très similaire au calage traditionnel. De toutes les analyses effectuées dans ce chapitre, seule la variabilité des paramètres vient distinguer ces deux stratégies. En effet, les paramètres des cinquante solutions obtenues par calage traditionnel avec ETR sont manifestement plus instables. Par contre, l'espace de recherche se voit diminué par l'ajout de la contrainte, d'où une légère diminution de la variabilité des solutions. Du reste, hydrogrammes et évapogrammes sont pratiquement superposés et la performance au

3. Certains travaux dépassant un peu le cadre de ce mémoire ont été menés en parallèle à la problèmatique principale, afin de valider ou d'infirmer certaines hypothèses ayant surgi en cours de route. Étant tout de même reliés à la thématique de ce projet, ils seront présentés en annexe. Le lecteur devra cependant noter que, ne faisant pas partie du mandat premier de ce mémoire, certaines pistes n'ont pas été investiguées en profondeur et seront reprises dans la Section 6.3 présentant les travaux futurs.

niveau des contraintes est équivalente. Rappelons cependant que les observations du bassin Caniapiscau étaient telles que le ratio $(1 - \frac{Q}{P})$ se trouvait déjà dans l'intervalle englobant la valeur de l'Atlas du Canada. Les résultats associés à d'autres bassins dont le ratio $(1 - \frac{Q}{P})$ est inférieur ou supérieur à la valeur de l'Atlas, présentés à l'Annexe G, seront donc des plus intéressants.

5.3.3 Calage parallèle

Finalement, le calage parallèle peut être comparé au calage traditionnel en termes de variabilité des paramètres. La découverte d'une forte corrélation entre la performance en calibration et en validation vient assurer la robustesse des solutions, mais aussi change la perspective des analyses, puisque seule la meilleure solution est souvent considérée. Ainsi, les résultats obtenus en calibration sont un peu dégradés par rapport aux autres stratégies, mais la meilleure solution en validation est comparable et la différence au niveau des hydrogrammes est difficile à détecter. Jusqu'à maintenant, le calage parallèle est équivalent, sans plus, aux autres stratégies, mis à part la corrélation des solutions qui offre un avantage certain. Lorsqu'on porte attention au processus d'évapotranspiration cependant, le doute n'est plus possible. La représentation du cycle annuel d'ETR telle que produite par le calage parallèle est beaucoup plus près de la réalité, représentée par la variable Qfs du modèle climatique, que toutes les représentations offertes par les autres stratégies de calage. Le plus marquant est probablement la période hivernale, où les stratégies de calage traditionnel et traditionnel avec bilan sur l'ETR ne parvienne même par à surpasser l'utilisation de la moyenne, alors que le calage parallèle permet un critère de Nash-Sutcliffe modifié $(1 - ns)$ autour de 0.7. L'optimisation du modèle avec cette stratégie de calage a permis de découvrir qu'une paramétrisation offrant à la fois une bonne représentation des débits et du processus d'évapotranspiration existe, et, de toutes les méthodes, seul le calage parallèle a réussi à la trouver.

5.3.4 Stratégie proposée en contexte de changements climatiques

Dans un contexte de changements climatiques, donc, où les objectifs visés par une paramétrisation sont bien différents que dans un contexte de prévision à court terme par exemple, la stratégie de calage parallèle est recommandée en regard des résultats obtenus dans le cadre de cette recherche. La meilleure des cinquante solutions trouvées permet non seulement une projection satisfaisante des apports naturels, mais aussi une représentation de la physique des processus sous la forme d'une reproduction réaliste du cycle de l'évapotranspiration réelle.

C'est donc une paramétrisation robuste et respectant la physique du processus d'évapotranspiration qui sera transposée dans un avenir hypothétique afin de permettre une estimation la plus fidèle possible des changements à prévoir.

5.4 Applications en changements climatiques

L'applicabilité en changements climatiques a aussi été sommairement validée sur le bassin versant Caniapiscau. Des tests plus exhaustifs seront réalisés dans les mois suivants le dépôt de ce mémoire et ceux-ci sont détaillés dans la Section 6.3. À partir des données météorologiques du modèle climatique MRCC piloté par trois différents scénarios de changements climatiques, il a été possible de tester les trois stratégies de calage afin de voir leur comportement en contexte de changements climatiques.

5.4.1 Utilisation du MRCC

La même version du MRCC a été utilisée pour cette section sur le même domaine centré sur le Québec (111x87 points de grille) avec résolution horizontale de 45 km. La différence réside dans les données utilisées pour le pilotage du modèle. Ainsi, les deux premières simulations, *aet* et *aev*, ont été pilotées à leurs frontières par les champs atmosphériques provenant d'une simulation du modèle canadien couplé de climat global (MCCG3/T47 à environ 3,75 x 3,75 deg. latitude-longitude ; 4è membre ; Scinocca *et al.* (2008) ; Flato *et al.* (2000), Flato et Boer (2001)). Les simulations globale et régionale ont été produites en suivant les scénarios projetés de gaz à effet de serre et d'aérosols du GIEC SRES A2 (Nakicenovic *et al.* (2000)). La troisième simulation quant à elle, *agx*, a été pilotée à ses frontières par les champs atmosphériques provenant d'une simulation du modèle allemand couplé de climat global (ECHAM5 à environ 1,87 deg latitude-longitude ; 1er membre ; Jungclaus *et al.* (2006)). Les simulations globale et régionale ont été produites en suivant les mêmes scénarios projetés de gaz à effet de serre et d'aérosols. Ces trois scénarios seront désignés respectivement par CGCM3.1#4, CGCM3.1#5 et ECHAM5. Deux périodes ont été simulées, soit la période dite de référence (REF) allant de 1961 à 2000 et la période future (FUT) de 2040 à 2069.

5.4.2 Résultats et analyses

Analyse sommaire des hydrogrammes

Les hydrogrammes obtenus par les cinquante optimisations de chacune des trois stratégies de calage, pour les trois scénarios climatiques, et pour les périodes de référence et future, sont présentés à la Figure 5.22. La stratégie de calage traditionnel ne semble pas être représentée, mais elle se trouve à être confondue et sous la stratégie de calage traditionnel avec bilan sur l'ETR et possède sensiblement la même dispersion. Puisque ce sont des scénarios climatiques présentant une réalité hypothétique, aucune référence n'est disponible afin de comparer la représentation des débits. Il est tout de même possible de constater une plus grande dispersion des solutions pour les calages traditionnel et traditionnel avec bilan sur l'ETR. Ceux-ci présentent aussi une montée de crue printanière plus hâtive et une crue d'automne plus tardive, encore qu'on ne puisse pas statuer sur la qualité de ces informations.

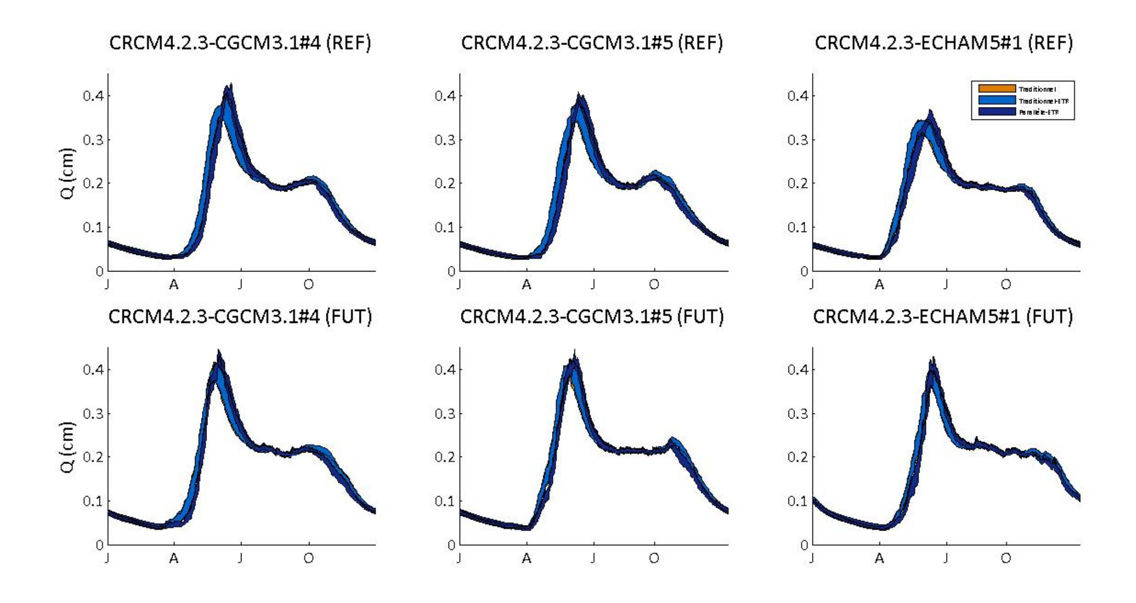


Figure 5.22 Hydrogrammes des trois stratégies de calage pour trois scénarios de changements climatiques, produit pour la période de référence et la période future.

Analyse sommaire des évapogrammes

Les évapogrammes correspondant aux hydrogrammes précédents apparaissent à la Figure 5.23. Comme pour les hydrogrammes, les courbes de la stratégie de calage traditionnel se trouvent sous celles du calage traditionnel avec bilan sur l'ETR et possèdent encore sensi-

blement la même dispersion. La première constatation à faire est que les courbes de ces deux stratégies de calage sont beaucoup plus dispersées que celles du calage parallèle. En période de référence, cette dernière stratégie permet une représentation du cycle de l'évapotranspiration plus juste que les stratégies de calage traditionnel et traditionnel avec ETR, surtout en période hivernale. L'ETR en été est cependant sous-estimée par toutes les stratégies et pour tous les scénarios.

Pour la période future, l'ensemble des stratégies, et particulièrement le calage parallèle, présente un cycle déphasé par rapport à celui généré par le modèle climatique. Ce décalage de la pointe observé sur l'évapogramme horizon 2050 peut s'expliquer notamment par une mauvaise séparation des précipitations entre la neige et la pluie. Cette hypothèse est apparue à quelques reprises dans le cadre de cette recherche, mais l'écart semble encore plus prononcé en période future. Une comparaison de la neige simulée par HSAMI à celle produite par le MRCC a donc été faite (Figure 5.24) afin de mieux comprendre le comportement du modèle hydrologique et cette hypothèse s'est vue confirmée, l'écart entre les deux modélisations de la neige étant beaucoup plus marqué pour l'horizon 2050. Notons que cette piste selon laquelle la séparation des phases serait problématique a été relevée à maintes reprises durant ce projet et sera explorée à l'Annexe G.2.

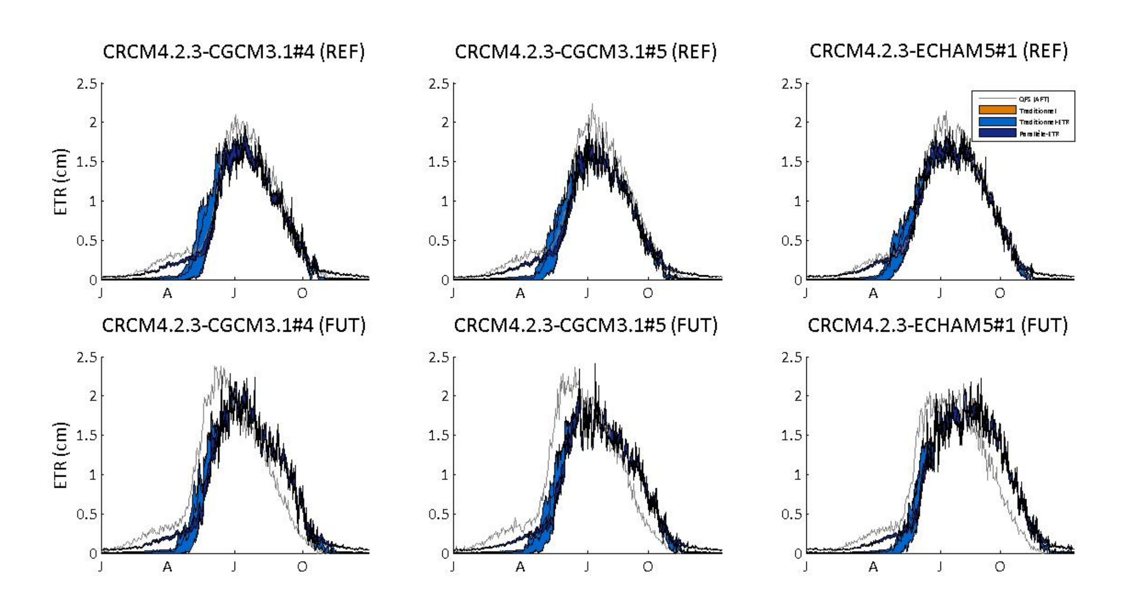


Figure 5.23 Évapogrammes des trois stratégies de calage pour trois scénarios de changements climatiques, produit pour la période de référence et la période future.

Changements d'ETR

En comparant le volume d'eau évapotranspiré sur la période de référence et sur la période future, il est possible d'arriver à un pourcentage de changement attendu dans l'évapotranspiration pour le futur. Un grand problème pour l'applicabilité des modèles hydrologiques en contexte de changements climatiques est l'énorme différence entre les modèles climatiques et hydrologiques pour ce changement. Comme il a été mentionné auparavant, les modèles climatiques peuvent estimer un changement jusqu'à trois fois plus important que les modèles hydrologiques (voir l'Annexe A). Puisque la valeur de ce changement a une grande importance dans la projection des débits en contexte de changements climatiques, un des objectifs des méthodes développées dans ce mémoire était de réduire l'écart avec les modèles climatiques.

La Figure 5.25 montre que les stratégies de calage traditionnel avec et sans contrainte sur l'ETR affiche un écart variant entre 2 % et 5 % par rapport à la stratégie de calage parallèle, qui est invariablement plus près du changement projeté par l'ensemble des scénarios climatiques. Plus exactement, la surestimation du changement en évapotranspiration est de 10 %, 14 % et 9 %, respectivement, pour chacun des scénarios, pour les stratégies de calage traditionnel et traditionnel avec ETR, et de 7 %, 9 % et 7 % pour le calage parallèle. Encore une fois, les stratégies de calage traditionnel et de calage traditionnel avec ETR offrent des résultats similaires, mais les solutions obtenues par la deuxième stratégie montrent une plus grande dispersion.

5.4.3 Conclusions

Pour l'ensemble des scénarios climatiques, la stratégie de calage parallèle propose des hydrogrammes et des évapogrammes moins dispersés que les autres stratégies. Il n'est pas possible de statuer sur la justesse des hydrogrammes, puisqu'il n'est pas possible de générer une série de débits de référence pour le futur. Par contre, pour les évapogrammes, la variable Qfs représentant l'évapotranspiration telle que modélisée par le modèle climatique, renseigne sur la rectitude des projections. Ainsi, la période de référence est mieux représentée par la stratégie de calage parallèle, surtout en période hivernale. Une sous-estimation de l'ETR en été résulte cependant de toutes les stratégies pour tous les scénarios. En période future, et de façon encore plus marquée pour le calage parallèle, tout le cycle d'évapotranspiration est déphasé par rapport à celui tracé à l'aide des valeurs de Qfs. Une explication impliquant la neige a tout de même été avancée et semble pouvoir expliquer ce phénomène. Finalement, l'écart avec le modèle climatique des changements projetés d'évapotranspiration en contexte

de changements climatiques est moins important pour la stratégie de calage parallèle. Il reste que le modèle hydrologique sous-estime encore le changement par rapport au MRCC. En somme, pour le bassin versant Caniapiscau, le calage parallèle semble encore une fois produire les résultats les plus justes.

5.5 Projet $(CQ)^2$

Plutôt que d'être reléguée au rang des bonnes idées qui, malheureusement sont tablettées, la stratégie de calage parallèle developpée au cours de ce projet a immédiatement été adoptée dans le cadre du projet $(CQ)^2$. Ce projet d'envergure, qui est une initiative conjointe de l'IREQ, du consortium Ouranos, du Centre d'expertise hydrique du Québec (CEHQ) et de l'entreprise Rio Tinto Alcan, a pour but principal de qualifier et de quantifier les impacts des changements climatiques sur le régime hydrologique québécois. Pour y arriver, une paramétrisation devra être obtenue pour chacun des 324 bassins versants recouvrant l'ensemble du territoire québécois (Figure 5.26), suivie d'une projection en changements climatiques sous 98 simulations climatiques employant 5 différentes méthodes de post-traitements. Au terme de ce projet, ce sont donc presque 160 000 scénarios hydroclimatiques qui auront été générés.

Le projet $(CQ)^2$ étant toujours en cours, les résultats ne seront pas présentés dans ce mémoire. Notons seulement que la paramétrisation des 324 bassins versants a été menée à terme et que les résultats obtenus semblent prometteurs. La stratégie de calage parallèle a d'ailleurs permis de détecter certaines inadéquations dans les données qui auraient pu passer inaperçues sous un calage traditionnel. La vaste base de données produite dans le cadre de ce projet offrira aussi les données nécessaires afin de juger des résultats obtenus par la méthode de calage parallèle sous différents scénarios de changements climatiques avec d'autres bassins versants que Caniapiscau. Nos conclusions quant à la généralistion de l'utilisation de cette stratégie pourront donc être étoffées.

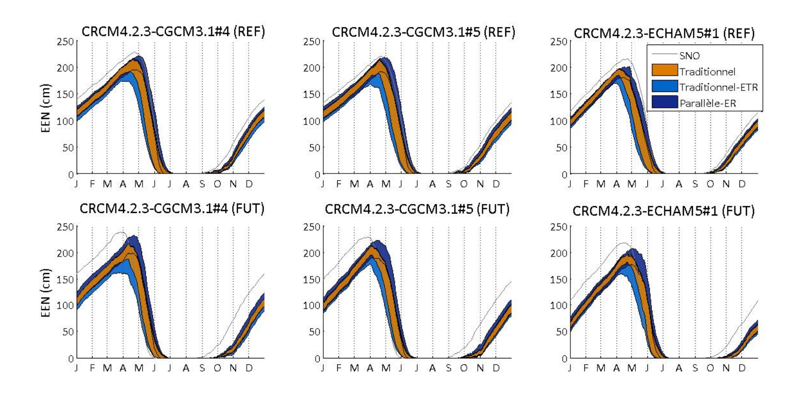


Figure 5.24 Confrontation de la variable SNO du MRCC et de la neige simulée par HSAMI.

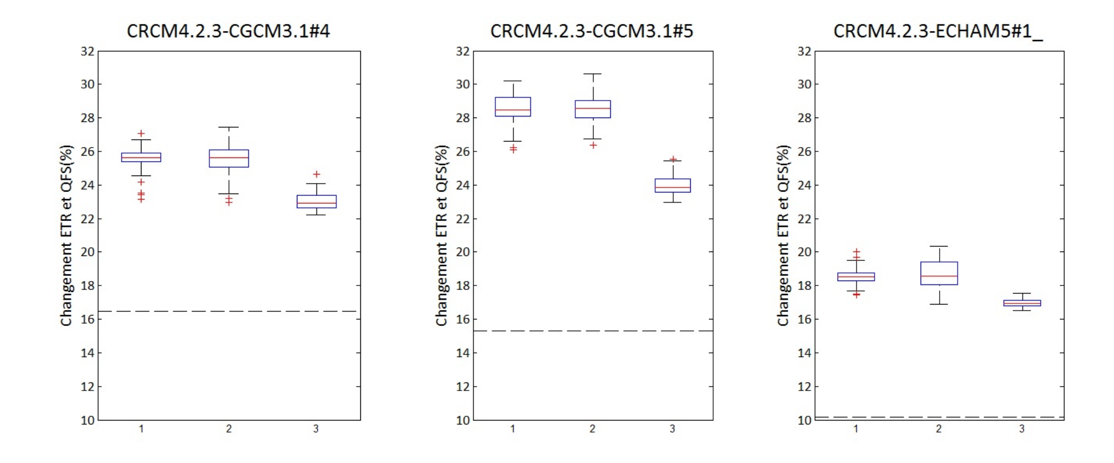


Figure 5.25 Changements de l'évapotranspiration prédits par cinquante simulations selon la stratégie de calage (traditionnel, traditionnel avec bilan sur l'ETR et parallèle sont respectivement 1, 2 et 3) et par le modèle climatique (en pointillé) pour trois scénarios de changements climatiques.



Figure 5.26 Carte des 324 bassins versants du projet $(CQ)^2$.

CHAPITRE 6

CONCLUSION

Les interrogations soulevées par le contexte futur incertain amène les entreprises dépendantes des ressources naturelles telles qu'Hydro-Québec à vouloir quantifier les impacts possibles du réchauffement planétaire sur leurs activités. Par le biais de modèles hydrologiques, des prévisions relativement fiables peuvent être produites à court et moyen termes. Ces modèles ne sont cependant pas conçus pour une utilisation en contexte de changements climatiques et c'est afin de répondre à cette demande de projections robustes que deux stratégies de calage visant à orienter le modèle vers la physique des processus ont été proposées. De ces deux méthodes, le calage parallèle mettant de l'avant la modélisation de l'évapotranspiration par un modèle climatique a su prouver la robustesse de ses solutions en combinant une bonne représentation à la fois des débits et du processus de l'évapotranspiration.

6.1 Synthèse des travaux

La principale contribution de ce mémoire au domaine de l'hydrologie est la présentation d'une méthode de calage robuste pouvant être transposée en contexte de changements climatiques et pouvant être étendue à plusieurs autres processus hydrologiques pour une représentation encore plus orientée vers la physique. Cette méthode peut non seulement s'appliquer à tout modèle global, mais peut aussi être adaptée aux modèles distribués (comme Hydrotel, utilisé chez Hydro-Québec) par l'orientation de certains autres processus hydrologiques comme la neige ou le débit de base. L'exploitation de cette méthode de calage n'en est donc qu'au commencement.

L'association avec le domaine de la recherche opérationnelle a aussi contribué à ce projet. Le regard nouveau porté sur les processus d'optimisation a permis une amélioration non négligeable de la performance de l'algorithme d'optimisation par le biais de l'emploi approprié des options offertes. Le lissage de la fonction a notamment permis à l'algorithme de trouver des résultats significativement meilleurs tout en réduisant le temps de calcul de 80 %. L'ajout de la VNS quant à elle a su engendrer des gains allant jusqu'à 12 % de la valeur de la fonction objectif. L'optimisation du modèle HSAMI sans cette fonctionnalité est maintenant officiellement déconseillée. L'utilisation du module complémentaire d'analyse de sensibilité a aussi

ouvert de nouvelles portes en offrant la possibilité de mettre à profit les connaissances des utilisateurs d'un modèle dans la formulation adéquate des contraintes. L'approche de calage par contraintes proposée dans cette recherche s'en est vu bonifiée et de nouvelles avenues de recherche en ont découlé. Finalement, la présentation des résultats par le biais des profils de performance est pratique courante dans le milieu de l'optimisation, mais n'était pas connue en hydrologie. Cette façon de faire représente une belle alternative aux diagrammes en boîtes qui ne permettaient pas toujours de faire ressortir adéquatement les différences entre les méthodes.

Pour le domaine de la recherche opérationnelle, ce projet a offert une belle opportunité d'application des outils de l'optimisation. L'algorithme MADS était déjà connu à l'IREQ suite au projet de Alarie *et al.* (2011) sur la localisation de senseurs à neige, mais son utilisation a été adaptée pour le calage de modèles hydrologiques, résultant en une amélioration de la qualité des solutions, mais aussi en une réduction du temps de calcul. Le module d'analyse de sensibilité quant à lui s'est fait connaître et a ouvert de nouvelles voies de recherche.

6.2 Limitations de la solution proposée

La première limitation de la méthode proposée dans ce mémoire est liée à l'utilisation de la paramétrisation obtenue suite à l'étalonnage du modèle. Le calage parallèle étant destiné à une étude des changements climatiques, l'exactitude des prévisions journalières est moins importante que l'ampleur des changements à prévoir entre la période de référence et la période future. Ce sont donc des critères dérivés de cette constatation qui ont servi dans l'analyse des performances de chacune des méthodes. Pour une application en prévisions à court terme, de nouvelles études devront nécessairement être réalisées.

Ensuite, les résultats obtenus dans cette recherche reposent sur nos hypothèses de base selon lesquelles le bilan de l'évapotranspiration réelle estimé par l'Atlas du Canada ainsi que la représentation de l'ETR telle que modélisée par le MRCC sont plus réalistes que la modélisation du processus par le modèle hydrologique HSAMI. La valeur fournie par l'Atlas du Canada est dérivée d'observations datant de 1978 et diffère de quelques unités de pourcentage selon la méthode d'interpolation utilisée. La représentation du cycle de l'ETR telle que modélisée par le modèle climatique, quant à elle, repose sur une seule simulation climatique soumise à un modèle régional, le MRCC. La validité de ces hypothèses pourrait être renforcée par la mise à jour de l'Atlas du Canada (projetée) et par la comparaison des cycles de l'ETR

produits par plusieurs autres modèles et simulations.

Finalement, les conclusions concernant les performances de chacune des trois stratégies de calage pourront être renforcées par une correction de la séparation des phases pluie-neige du modèle hydrologique. Le déficit en neige a posé plusieurs problèmes au cours de ce projet, notamment dans la reproduction adéquate des débits. La montée de crue s'est ainsi vue retardée de quelques semaines lorsque la contrainte d'évapotranspiration forçait le modèle à évaporer l'eau qui aurait normalement dûe se retrouver dans la neige. Une correction de la séparation des phases devrait corriger ce problème et il serait pertinent de le vérifier.

6.3 Améliorations futures

L'approche proposée dans ce mémoire étant à ses débuts, plusieurs pistes n'ont pu être explorées, faute de temps. Elles seront présentées ici et reprises dans des travaux futurs, notamment par Marie Minville. Ainsi, seule la fonction objectif couramment utilisée chez Hydro-Québec, soit le critère de Nash-Sutcliffe, a été appliquée. Cette fonction ayant ses forces et ses faiblesses, la recherche d'une alternative pourrait éventuellement être considérée. De la même façon, un certain nombre de bassins a dû être ciblé pour les études préliminaires. Cependant, comme il a été fait mention précédemment, le projet $(CQ)^2$ a permis de généraliser la méthode à l'ensemble du Québec.

Un seul processus hydrologique, l'évapotranspiration, a pu être contraint pour cette première phase de tests. Sachant que la méthode est prometteuse, des travaux pourront maintenant être dirigés sur d'autres processus, tels que la neige ou le débit de base. D'ailleurs, des tests préliminaires ont déjà été réalisés sur ce dernier et l'impact sur l'évapotranspiration est notoire. La stratégie de calage parallèle, de par sa flexibilité, permet aussi de contraindre le modèle à l'aide de plusieurs versions parallèles pilotées par différents scénarios climatiques ou sur plusieurs périodes. Les possibilités sont encore une fois presqu'illimitées en ce sens où chaque processus pourrait être contraint indépendamment, mais simultanément. La recherche sera définitivement poursuivie afin d'explorer ces alternatives.

Au niveau de l'optimisation, deux pistes de recherche ressortent de ce mémoire. En premier lieu, il pourrait être intéressant de tester plusieurs métaheuristiques du même type que la recherche à voisinage variable étant donné les résultats obtenus suite à son utilisation. Par exemple, l'IREQ a développé à l'interne sa propre version de la métaheuristique et il

serait pertinent de comparer sa performance à celle de la VNS de Mladenović et Hansen (1997). Dans le même ordre d'idées, seules certaines fonctionnalités de NOMAD ont été personnalisées, avec des impacts majeurs sur l'optimisation. L'étude des autres fonctionnalités disponibles avec NOMAD, mais aussi avec d'autres optimiseurs, pourrait résulter en un optimiseur en tout point pensé pour notre problème et, donc, menant potentiellement à de meilleures performances.

RÉFÉRENCES

ABRAMSON, M., AUDET, C., DENNIS, JR., J. et LE DIGABEL, S. (2009). OrthoMADS : A deterministic MADS instance with orthogonal directions. *SIAM Journal on Optimization*, 20, 948–966.

ALARIE, S., AUDET, C., GARNIER, V., LE DIGABEL, S. et LECLAIRE, L. (2011). Snow water equivalent estimation using blackbox optimization. Rapport technique G-2011-09, Les cahiers du GERAD. To appear in *Pacific Journal of Optimization*.

ANCTIL, F., ROUSSELLE, J. et LAUZON, N. (2005). *Hydrologie - cheminements de l'eau*. Presses Internationales Polytechnique.

AUDET, C. (2004). Convergence results for pattern search algorithms are tight. *Optimization and Engineering*, 5, 101–122.

AUDET, C., BÉCHARD, V. et LE DIGABEL, S. (2008a). Nonsmooth optimization through mesh adaptive direct search and variable neighborhood search. *Journal of Global Optimization*, 41, 299–318.

AUDET, C., CUSTÓDIO, A. et DENNIS, JR., J. (2008b). Erratum : Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 18, 1501–1503.

AUDET, C. et DENNIS, JR., J. (2003). Analysis of generalized pattern searches. *SIAM Journal on Optimization*, 13, 889–903.

AUDET, C. et DENNIS, JR., J. (2004). A pattern search filter method for nonlinear programming without derivatives. *SIAM Journal on Optimization*, 14, 980–1010.

AUDET, C. et DENNIS, JR., J. (2006). Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 17, 188–217.

AUDET, C. et DENNIS, JR., J. (2009). A progressive barrier for derivative-free nonlinear programming. *SIAM Journal on Optimization*, 20, 445–472.

AUDET, C., DENNIS, JR., J. et LE DIGABEL, S. (2008c). Parallel space decomposition of the mesh adaptive direct search algorithm. *SIAM Journal on Optimization*, 19, 1150–1170.

AUDET, C., DENNIS, JR., J. et LE DIGABEL, S. (2010a). Globalization strategies for mesh adaptive direct search. *Computational Optimization and Applications*, 46, 193–215.

AUDET, C., DENNIS, JR., J. et LE DIGABEL, S. (2012). Trade-off studies in blackbox optimization. *Optimization Methods and Software*, 27, 613–624.

AUDET, C., SAVARD, G. et ZGHAL, W. (2008d). Multiobjective optimization through a series of single-objective formulations. *SIAM Journal on Optimization*, 19, 188–210.

AUDET, C., SAVARD, G. et ZGHAL, W. (2010b). A mesh adaptive direct search algorithm for multiobjective optimization. *European Journal of Operational Research*, 204, 545–556.

AWAD, A. R. et VON POSER, I. (2010). Calibrating conceptual rainfall-runoff models using a real genetic algorithm combined with a local search method. *LATEST TRENDS on COMPUTERS (volume I), Proceedings of the 14th WSEAS international*. 174 – 181.

BERGSTRÖM, S., LINDSTRÖM, G. et PETTERSSON, A. (2002). Multi-variable parameter estimation to increase confidence in hydrological modelling. *Hydrological Processes*, 16, 413–421.

BEVEN, K. (1996). Equifinality and uncertainty in geomorphological modelling. *The Scientific Nature of Geomorphology : Proceedings of the 27th Binghamton Symposium in Geomorphology held 27-29 September 1996*.

BEVEN, K. (2001). On hypothesis testing in hydrology. *Hydrological Processes*, 15, 1655–1657.

BEVEN, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320, 18 – 36. <ce :title>The model parameter estimation experiment</ce :title> <ce :subtitle>MOPEX</ce :subtitle> <xocs :full-name>MOPEX workshop</xocs :full-name>.

BILLUPS, S., DIRKSE, S. et FERRIS, M. (1997). A comparison of algorithms for large-scale mixed complementarity problems. *Comput. Optim. Appl.*, 7, 3 – 25.

BINDLISH, R. et BARROS, A. P. (2000). Disaggregation of rainfall for one-way coupling of atmospheric and hydrologic models in regions of complex terrain. *Global Planet. Change*, 25, 111–132.

BISSON, J. et ROBERGE, F. (1983). Prévision des apports naturels : Expérience d'hydro-québec. *Compte-rendu de l'Atelier sur la prévision du débit*.

BOOKER, A., DENNIS, JR., J., FRANK, P., SERAFINI, D., TORCZON, V. et TROSSET, M. (1999). A rigorous framework for optimization of expensive functions by surrogates. *Structural and Multidisciplinary Optimization*, 17, 1–13.

CARRER, G., ROUSSEAU, A. N. et RIVARD, C. (2009). Effets des changements climatiques sur la recharge de bassins versants de l'est du canada. Rapport technique R-1049, Centre Eau Terre et Environnement de l'Institut national de la recherche scientifique (INRS-ETE).

CAYA, D. et LAPRISE, R. (1999). A semi-implicit semi-lagrangian regional climate model : The canadian rcm. *Monthly Weather Review*, 127, 341–362.

CHIEW, F. (1993). Comparison of six rainfall-runoff modelling approaches. *Journal of Hydrology*, 147, 1–36.

CHIEW, F., TENG, J., VAZE, J. et KIRONO, D. (2009). Influence of global climate model selection on runoff impact assessment. *Journal of Hydrology*, 379, 172 – 180.

CHKIR, N. (1994). *Mise au point d'un modèle hydrologique conceptuel intégrant l'état hydrique du sol dans la modélisation pluie-débit*. Mémoire de maîtrise, École Nationale des Ponts et Chaussées de Paris.

CHOCAT, B. (1997). *Encyclopédie de l'hydrologie urbaine et de l'assainissement*.

CLARKE, F. (1983). *Optimization and Nonsmooth Analysis*. John Wiley & Sons, New York. Réedité en 1990 par SIAM Publications, Philadelphia, comme le Vol. 5 de la série Classics in Applied Mathematics.

CONN, A. et LE DIGABEL, S. (2011). Use of quadratic models with mesh adaptive direct search for constrained black box optimization. Rapport technique G-2011-11, Les cahiers du GERAD. To appear in *Optimization Methods and Software*.

COOPER, V., NGUYEN, V.-T.-V. et NICELL, J. (2007). Calibration of conceptual rainfall-runoff models using global optimisation methods with hydrologic process-based parameter constraints. *Journal of Hydrology*, 334, 455 – 466.

DANTZIG, G. (1951). Maximization of a linear function of variables subject to linear inequalities : Activity analysis of production and allocation, 339–347.

DAVIDON, W. (1991). Variable metric method for minimization. *SIAM Journal on Optimization*, 1.

DEB, K., PRATAP, A., AGARWAL, S. et MEYARIVAN, T. (2002). A fast and elitist multiobjective genetic algorithm : Nsga-ii. *IEEE Transactions on evolutionary computation*. vol. 6, 182–197.

DESJARLAIS, C. *ET AL.* (2010). S'adapter aux changements climatiques. Rapport technique, Consortium Ouranos.

DOLAN, E. et MORÉ, J. (2002). Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91, 201–213.

DUAN, Q., SOROOSHIAN, S. et GUPTA, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28, 1015 – 1031.

EFSTRATIADIS, A. et KOUTSOYIANNIS, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling : a review. *Hydrological Sciences Journal*, 55, 58–78.

FLATO, G. et BOER, G. (2001). Warming asymmetry in climate change simulations. *Geophys. Res. Lett.*, 28, 195–198.

FLATO, G., BOER, G., LEE, W., MCFARLANE, N., RAMSDEN, D., READER, M. et WEAVER, A. (2000). The canadian centre for climate modeling and analysis global coupled model and its climate. *Climate Dynamics*, 16, 451–467.

FLETCHER, R. et LEYFFER, S. (2002). Nonlinear programming without a penalty function. *Mathematical Programming*, Series A, 91, 239–269.

FORTIN, V. (1999). Le modèle météo-apport hsami : historique, théorie et application. Rapport technique Rapport IREQ-1999-0255, Institut de recherche d'Hydro-Québec.

FRANCHINI, M. (1991). Comparative analysis of several conceptual rainfall-runoff models. *Journal of Hydrology*, 122, 161–219.

GAN, T. Y., DLAMINI, E. M. et BIFTU, G. F. (1997). Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *Journal of Hydrology*, 192, 81 – 103.

GARRICK, M., CUNNANE, C. et NASH, J. (1978). A criterion of efficiency for rainfall-runoff models. *Journal of Hydrology*, 36, 375 – 381.

GNOUMA, R. (2006). Aide à la calibration d'un modèle hydrologique distribué au moyen d'une analyse des processus hydrologiques : application au bassin versant de l'yzeron.

GOSWAMI, M. et O'CONNOR, K. M. (2007). Comparative assessment of six automatic optimization techniques for calibration of a conceptual rainfall-runoff model. *Hydrological Sciences Journal*, 52, 432–449.

GRAMACY, R. et LE DIGABEL, S. (2011). The mesh adaptive direct search algorithm with treed Gaussian process surrogates. Rapport technique G-2011-37, Les cahiers du GERAD.

GUPTA, H., SOROOSHIAN, S. et YAPO, P. (1999). Status of automatic calibration for hydrologic models : Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, 4, 135 – 143.

GUPTA, H. V., SOROOSHIAN, S. et YAPO, P. O. (1996). Automatic calibration of conceptual rainfall-runoff models : sensitivity to calibration data. *Journal of Hydrology*, 181, 23 – 48.

GUPTA, H. V., SOROOSHIAN, S. et YAPO, P. O. (1998). Toward improved calibration of hydrologic models : Multiple and noncommensurable measures of information. *Water Resources research*, 34, 751 – 763.

HANSEN, P. et MLADENOVIĆ, N. (2001). Variable neighborhood search : principles and applications. *European Journal of Operational Research*, 130, 449–467.

JUNGCLAUS, J., BOTZET, M., HAAK, H., KEENLYSIDE, N., LUO, J.-J., LATIF, M., MAROTZKE, J., MIKOLAJEWICZ, U. et ROECKNER, E. (2006). Ocean circulation and tropical variability in the aogcm echam5/mpi-om. *J. Clim.*, 19, 3952–3972.

KHU, S. T. et MADSEN, H. (2005). Multiobjective calibration with pareto preference ordering : An application to rainfall-runoff model calibration. *Water Resources research*, 41.

KRALISCH, S., KRAUSE, P., FINK, M., C., F. et FLÜGEL, W.-A. (2007). Component based environmental modelling using the jams framework. L. Oxley et D. Kulasiri, éditeurs, *MODSIM 2007 International Congress on Modelling and Simulation*. Modelling and Simulation Society of Australia and New Zealand, Christchurch, New Zealand, 812 – 818.

LAFORCE, S., SIMARD, M.-C., LECONTE, R. et BRISSETTE, F. (2011). Climate change and floodplain delineation in two southern quebec river basins. *JAWRA Journal of the American Water Resources Association*, 47, 785–799.

LE DIGABEL, S. (2011). Algorithm 909 : NOMAD : Nonlinear optimization with the MADS algorithm. *ACM Transactions on Mathematical Software*, 37, 44 :1–44 :15.

LE DIGABEL, S. et TRIBES, C. (2009). NOMAD user guide. Rapport technique G-2009-37, Les cahiers du GERAD.

LECLAIRE, L.-A. (2011). Communication personnelle.

LOFGREN, B., HUNTER, T. et WILBARGER, J. (2011). Effects of using air temperature as a proxy for potential evapotranspiration in climate change scenarios of great lakes basin hydrology. *Journal of Great Lakes Research*.

MADSEN, H. (2000). Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology*, 235, 276 – 288.

MATHIEU, C. (2003). Rapport d'étalonnage : Projet sur l'impact des changements climatiques sur le débit de sortie à la centrale de carillon. Rapport technique, Hydro-Québec Production.

MILLY, P. et DUNNE, K. A. (2011). On the hydrologic adjustement of climate-model projections : The potential pitfall of potential evapotranspiration. *Earth Interactions*, 15, 1 – 14.

MIMIKOU, M., BALTAS, E., VARANOU, E. *ET AL.* (2000). Regional impacts of climate change on water resources quantity and quality indicators. *Journal of hydrology*, 234, 95–109.

MINVILLE, M. (2012). Recommandation d'un ensemble de données météorologiques pour l'exercice capacité de produire en contexte de changements climatiques. Rapport technique Rapport IREQ-2012-0036, Institut de recherche d'Hydro-Québec.

MINVILLE, M., GUAY, C. et LECLAIRE, L.-A. (2012). Nouvelles approches de calage du modèle hydrologique hsami pour une paramétrisation orientée vers la physique et applications en contextes actuel et futur. Rapport technique, Institut de recherche d'Hydro-Québec.

MINVILLE, M., KRAU, S., BRISSETTE, F. et LECONTE, R. (2010). Behaviour and performance of a water resource system in québec (canada) under adapted operating policies in a climate change context. *Water resources management*, 24, 1333–1352.

MLADENOVIĆ, N. et HANSEN, P. (1997). Variable neighborhood search. *Computers and Operations Research*, 24, 1097–1100.

MORÉ, J. et WILD, S. (2009). Benchmarking derivative-free optimization algorithms. *SIAM Journal on Optimization*, 20, 172–191.

MUSIC, B. et CAYA, D. (2007). Evaluation of the hydrological cycle over the mississippi river basin as simulated by the canadian regional climate model (crcm). *J. Hydrometeorology*, 8, 969–988.

NAKICENOVIC, N., SWART, S. *ET AL.* (2000). Ipcc special report on emissions scenarios : a special report of working group iii of the ipcc. Rapport technique, IPCC. 599 pages.

PAPPENBERGER, F. et BUIZZA, R. (2009). The skill of ecmwf precipitation and temperature predictions in the danube basin as forcings of hydrological models. *Weather and Forecasting*, 24, 749 – 766.

PIETRONIRO, A., SOULIS, E. D., SNELGROVE, K. et KOUWEN, N. (2001). A framework for coupling atmospheric and hydrological models. *Soil-Vegetation-Atmosphere Transfer Schemes and Large-Scale Hydrological Models.* Proceedings of a symposium held during the Sixth IAHS Scientific Assembly at Maastricht, The Netherlands, IAIIS Publ. no. 270.

PRUDHOMME, C., DÖRTE, J. et SVENSSON, C. (2003). Uncertainty and climate change impact on the flood regime of small uk cathments. *Journal of hydrology*, 277, 1–23.

REFSGAARD, J. et STORM, B. (1996). Construction, calibration and validation of hydrological models. M. B. Abbott et J. Refsgaard, éditeurs, *Distributed hydrological modelling.* 41 – 54.

ROBERT, B., PELLERIN, P. *ET AL.* (2000). Toward the use of coupled atmospheric and hydrologic models at regional scale. *Monthly Weather Review*, 128, 1681–1706.

ROY, J. (2011). *Étude d'impact de l'optimisation des règles de gestion en condition de changements climatiques sur le système hydrique de la rivière Manicouagan.* Mémoire de maîtrise, École de technologie supérieure.

SAID, M. (2006). Identification et prise en compte des sources d'incertitude dans les modèles conceptuels météo-apport. Rapport technique, Centre Eau Terre et Environnement de l'Institut national de la recherche scientifique (INRS-ETE).

SALATHÉ JR, E. P., MOTE, P. W. et WILEY, M. W. (2007). Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the united states pacific northwest. *International Journal of Climatology*, 27, 1611–1621.

SCHEAFLI, B. et GUPTA, H. (2007). Do nash values have value ? *Hydrological Processes*, 21, 2075–2080.

SCINOCCA, J. F., MCFARLANE, N. A., LAZARE, M., LI, J. et PLUMMER, D. (2008). Technical note : The cccma third generation agcm and its extension into the middle atmosphere. *Atmos. Chem. Phys.*, 8, 7055–7074.

SINGH, V. (1995). *Computer models of watershed hydrology.* Water Resources Publications. Ill., maps. bibliographical ref. and index.

SOULIS, E. D., SNELGROVE, K. R., KOUWEN, N. et SEGLENIEKS, F. (2000). Towards closing the vertical water balance in canadian atmospheric models : Coupling of the land surface scheme class with the distributed hydrological model watflood. *Atmosphere-Ocean*, 38, 251–269.

TORCZON, V. (1997). On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 7, 1–25.

UPPALA, S. *ET AL.* (2005). The era-40 re-analysis. *Q.J.R. Meteorol. Soc.*, 131, 2961–3012.

VICENTE, L. et CUSTÓDIO, A. (2012). Analysis of direct searches for discontinuous functions. *Mathematical Programming*, 133, 299–325.

YAHIA, I. B. (2012). *Identification statistique de variables importantes pour l'optimisation de boîtes noires.* Mémoire de maîtrise, École Polytechnique de Montréal.

YANG et LIU (2009). Automatic calibration of a rainfall-runoff model using a fast and elitist multi-objective particle swarm algorithm. *Expert Systems with Applications*, 36, 9533–9538.

ANNEXE A

CHANGEMENTS D'ÉVAPOTRANSPIRATION

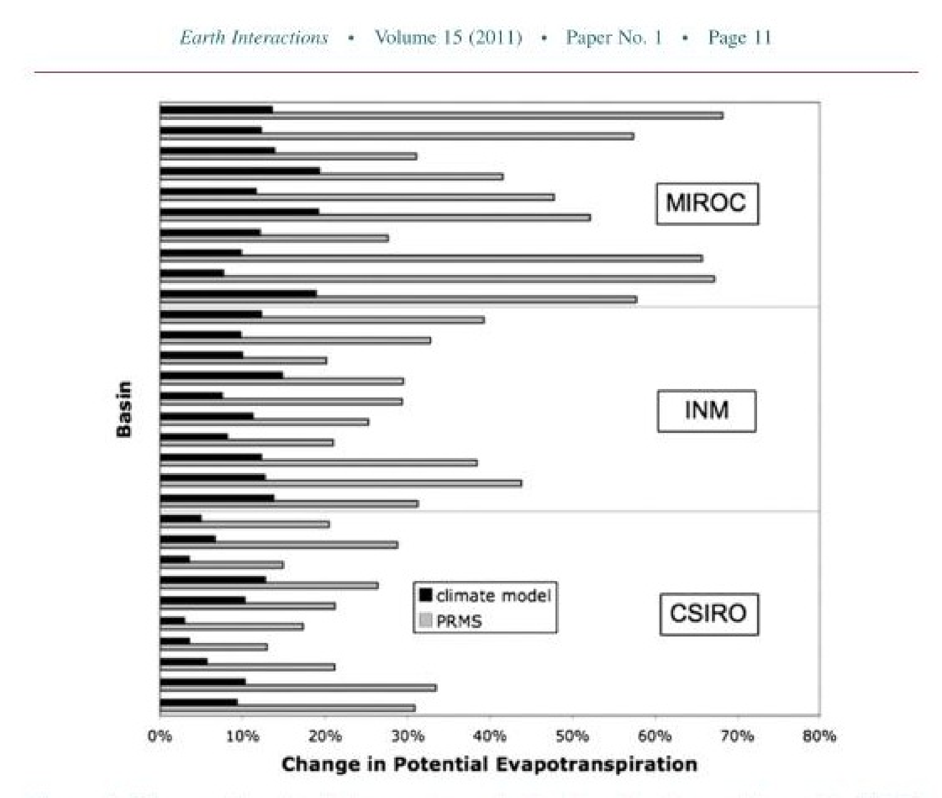


Figure 3. Changes in potential evapotranspiration for climate models and for PRMS for the time period 2088–99 under the A1B scenario, expressed relative to the baseline period 1988–99. Three sections of the chart correspond to the three climate models from which climate-change information is obtained. Within each section, basins are ordered alphabetically from the bottom of the chart upward by short name (Table 1).

Figure A.1 Changements d'ETR : modèles climatiques versus modèles hydrologiques (Milly et Dunne (2011)).

ANNEXE B

LE MODÈLE HSAMI

Intrants, paramètres et variables internes du modèle HSAMI, tirés du guide de l'utilisateur produit par Fortin (1999)

Tableau B.1 Intrants du modèle HSAMI.

Équivalents			Variables météorologiques utilisées comme intrant pour le modèle HSAMI (note: il s'agit toujours de valeurs moyennes pour le bassin versant)		
PL/1	C++		variable météorologique		signification
5/9 * (MIN-32)	5/9 * (TempMin-32)	1	dans ce rapport: dans le code *MATLAB*:	T_{min} (C) t_min (C)	température minimale quotidienne (quel que soit le pas de temps de simulation)
5/9 * (MAX-32)	5/9 * (TempMax-32)	2	dans ce rapport: dans le code *MATLAB*:	T_{max} (C) t_max (C)	température maximale quotidienne (quel que soit le pas de temps de simulation)
2,54 * PLUIE	2,54 * Pluie	3	pluie (cm)		quantité de pluie tombée au cours du pas de temps
2,54 * NEIGE	2,54 * Neige	4	neige (cm)		équivalent en eau de la neige tombée durant le pas de temps
ENSOL	Soleil	5	soleil (entre 0 et 1)		indice d'ensoleillement quotidien
	2,54 * NeigeAuSol	6	neige_au_sol (cm)		équivalent en eau du stock de neige au sol (optionnel)

Tableau B.2 Paramètres du modèle HSAMI.

Équivalent PL/1 ou C++		paramètres liés à l'évapotranspiration	rôle du paramètre
FAETP	1	efficacite_evapo_ete	facteur multipliant l'ETP permettant d'estimer l'ETR en été
POUR	2	efficacite_evapo_hiver	facteur multipliant *efficacite_evapo_ete**ETP permettant d'estimer l'ETR en hiver
		paramètres liés à la fonte de neige	**rôle du paramètre**
12,7 / 9 * FKM	3	taux_fonte_jour (cm/ΔC/jour)	taux de fonte du couvert nival, en cm/jour, pour chaque degré Celcius d'écart entre la température maximale et le seuil *temp_fonte_jour*
12,7 / 9 * FKYN	4	taux_fonte_nuit (cm/ΔC/jour)	taux de fonte du couvert nival, en cm/jour, pour chaque degré Celcius d'écart entre la température minimale et le seuil *temp_fonte_nuit*
5/9 * (TB-32)	5	temp_fonte_jour (C)	seuil de température maximale à partir duquel le couvert de neige peut fondre
5/9 * (TBN-32)	6	temp_fonte_nuit (C)	seuil de température minimale à partir duquel la fonte du couvert de neige est accélérée
5/9 * (TACAL-32)	7	temp_ref_pluie (C)	température de référence par rapport à laquelle on calcule la chaleur apportée par la pluie au couvert nival
CC	8	effet_redoux_sur_aire_enneigee	paramètre empirique utilisé pour relier les variables d'état *neige_au_sol_totale* et *fonte_totale* à la proportion du bassin couvert par de la neige
		paramètres liés à l'estimation du ruissellement de surface	**rôle du paramètre**
CGEL	9	effet_gel	paramètre empirique utilisé pour relier l'écart entre les variables d'état *gel* et *degel* à la proportion de l'eau de fonte qui ruisselle en surface
2,54 * CIMAX	10	effet_sol (cm)	quantité de pluie sur 24h nécessaire pour que 50% de cette eau ruisselle en surface lorsque le sol est complètement sec (i.e. la réserve *sol* est égale à zéro)
2,54 * CSAP	11	seuil_min (cm)	quantité de pluie sur 24h nécessaire pour que 50% de cette eau ruisselle en surface lorsque le sol est saturé (i.e. la réserve *sol* est égale à *sol_max*)

Tableau B.3 Paramètres du modèle HSAMI (suite).

Équivalent PL/1 ou C++	paramètres liés à l'écoulement vertical de l'eau dans le sol	rôle du paramètre
2,54 * RMIN	12 sol_min (cm)	quantité d'eau dans la zone non-saturée qui ne peut s'écouler par gravité
2,54 * RMAX	13 sol_max (cm)	quantité maximale d'eau que peut contenir la zone non-saturée du sol
2,54 * SAPMAX	14 nappe_max (cm)	quantité maximale d'eau que peut contenir l'aquifère avant de générer un ruissellement de surface
D	15 portion_ruissellement_surface	proportion de l'eau en surface qui ruisselle via l'hydrogramme intermédiaire au lieu de cheminer à travers la colonne de sol
X	16 portion_ruissellement_sol_max	proportion de l'eau qui ruisselle via l'hydrogramme intermédiaire lorsque la réserve non-saturée déborde (le reste est acheminé vers la nappe)
PER	17 taux_vidange_sol_min ($24h^{-1}$)	taux de vidange de la zone non-saturée vers la nappe
Y	18 taux_vidange_nappe ($24h^{-1}$)	taux de vidange de la nappe, cette vidange constituant le débit de base
	paramètres liés à l'écoulement horizontal de l'eau vers l'exutoire	**rôle du paramètre**
ALP	19 taux_vidange_inter ($24h^{-1}$)	taux de vidange de la réserve intermédiaire, cette vidange constituant l'apport intermédiaire
C++: TPOINT PL/1:ALPHA/BETA	20 mode_hydrogramme_surface (jour)	pointe de l'hydrogramme unitaire de surface
BETA	21 forme_hydrogramme_surface	paramètre de forme de l'hydrogramme de surface (selon une fonction de distribution gamma)
BET/GAMMA	22 mode_hydrogramme_inter (jour)	pointe de l'hydrogramme intermédiaire
GAMMA	23 forme_hydrogramme_inter	paramètre de forme de l'hydrogramme de surface (selon une fonction de distribution gamma)

Tableau B.4 Variables internes du modèle HSAMI.

Équivalent PL/1 ou C++	variable d'état et unité de mesure	rôle de la variable
2,54 * SNEI	1 neige_au_sol (cm)	équivalent en eau du couvert nival, incluant l'eau libre dans la neige
2,54 * F	2 fonte (cm)	quantité d'eau libre dans la neige
2,54 * SNEIM	3 neige_au_sol_totale (cm)[1]	équivalent en eau de la quantité totale de neige tombée depuis l'apparition du couvert nival (utilisée uniquement pour estimer la proportion du bassin couvert par de la neige)
2,54 * FT	4 fonte_totale (cm)[1]	quantité totale de neige qui a fondue depuis l'apparition du couvert nival (utilisée uniquement pour estimer la proportion du bassin couvert par de la neige)
-FANEI/0,38	5 derniere_neige (jours)[1]	nombre de jours depuis la dernière chute de neige (permet d'estimer l'albédo de la neige)
2,54 * GEL	6 gel (cm)	quantité d'eau gelée dans le sol et en surface du sol
2,54 * DEGEL	7 degel (cm)[1,2]	indice de dégel du sol (utilisé uniquement pour évaluer l'effet du gel sur le ruissellement de surface)
2,54 * R	8 sol (cm)	quantité d'eau dans la zone non-saturée du sol
2,54 * SAP	9 nappe (cm)	quantité d'eau dans la zone saturée du sol (eau contribuant au débit de base)
2,54 * Q	10 reserve_inter (cm)	quantité d'eau dans la réserve intermédiaire (eau en transit vers l'exutoire par ruissellement retardé)

[1]ces variables d'état ne correspondent pas à des réservoirs du modèle conceptuel
[2]cette variable d'état n'est utilisée que dans la version 1 du modèle conceptuel

ANNEXE C

BORNES UTILISÉES

Tableau C.1 Bornes des paramètres du modèle HSAMI.

INF	Paramètre	SUP
0.90000	Paramètre 1	1.15000
0.00000	Paramètre 2	0.20000
0.04000	Paramètre 3	0.50000
0.04000	Paramètre 4	0.20000
-4.00000	Paramètre 5	4.00000
-4.00000	Paramètre 6	4.00000
-2.00000	Paramètre 7	2.00000
1.10000	Paramètre 8	3.00000
1.00000	Paramètre 9	10.00000
15.00000	Paramètre 10	35.00000
1.30000	Paramètre 11	6.00000
0.00000	Paramètre 12	10.00000
10.00000	Paramètre 13	20.00000
8.00000	Paramètre 14	23.00000
0.25000	Paramètre 15	0.55000
0.40000	Paramètre 16	0.90000
0.01000	Paramètre 17	0.04000
0.00800	Paramètre 18	0.01300
0.90000	Paramètre 19	0.97000
0.50000	Paramètre 20	3.00000
0.70000	Paramètre 21	3.00000
3.00000	Paramètre 22	10.00000
0.30000	Paramètre 23	2.00000

ANNEXE D

BORNES DE L'INRS

Tableau D.1 Bornes de l'Institut national de la recherche scientifique.

INF	Paramètre	SUP
0.00000	Paramètre 1	1.2000
0.00000	Paramètre 2	1.00000
0.00000	Paramètre 3	2.00000
0.00000	Paramètre 4	2.00000
-4.00000	Paramètre 5	4.00000
-4.00000	Paramètre 6	4.00000
-4.00000	Paramètre 7	4.00000
0.00000	Paramètre 8	10.00000
0.00000	Paramètre 9	10.00000
0.00000	Paramètre 10	100.00000
0.00000	Paramètre 11	10.00000
0.00000	Paramètre 12	10.00000
0.00000	Paramètre 13	20.00000
0.00000	Paramètre 14	20.00000
0.00000	Paramètre 15	1.00000
0.00000	Paramètre 16	1.00000
0.00000	Paramètre 17	1.00000
0.00000	Paramètre 18	1.00000
0.00000	Paramètre 19	1.00000
0.00000	Paramètre 20	5.00000
0.00000	Paramètre 21	5.00000
0.00000	Paramètre 22	10.00000
0.00000	Paramètre 23	5.00000

ANNEXE E

CARTES DE L'ATLAS DU CANADA

Figures dérivées de la carte du bilan hydrique de l'Atlas du Canada provenant du site web de Ressources naturelles Canada

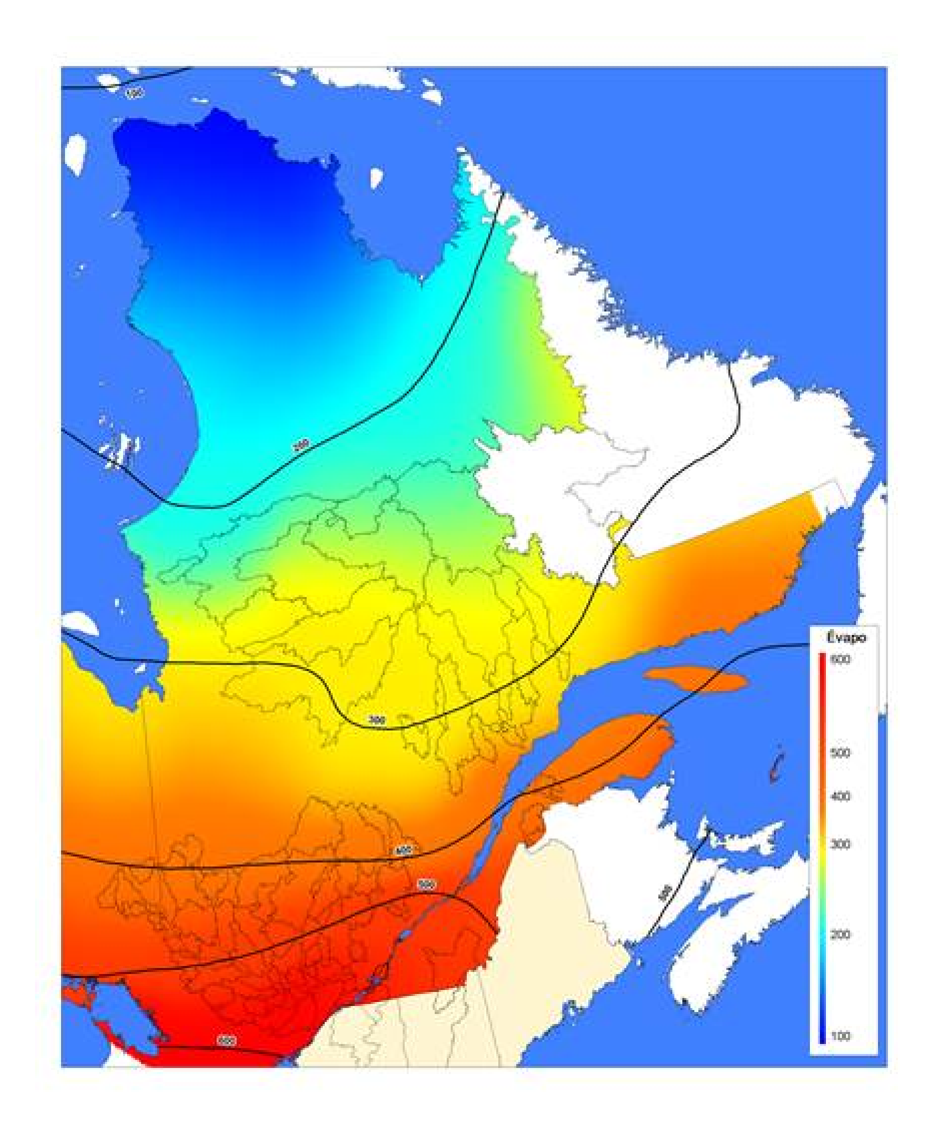


Figure E.1 Évapotranspiration dérivée.

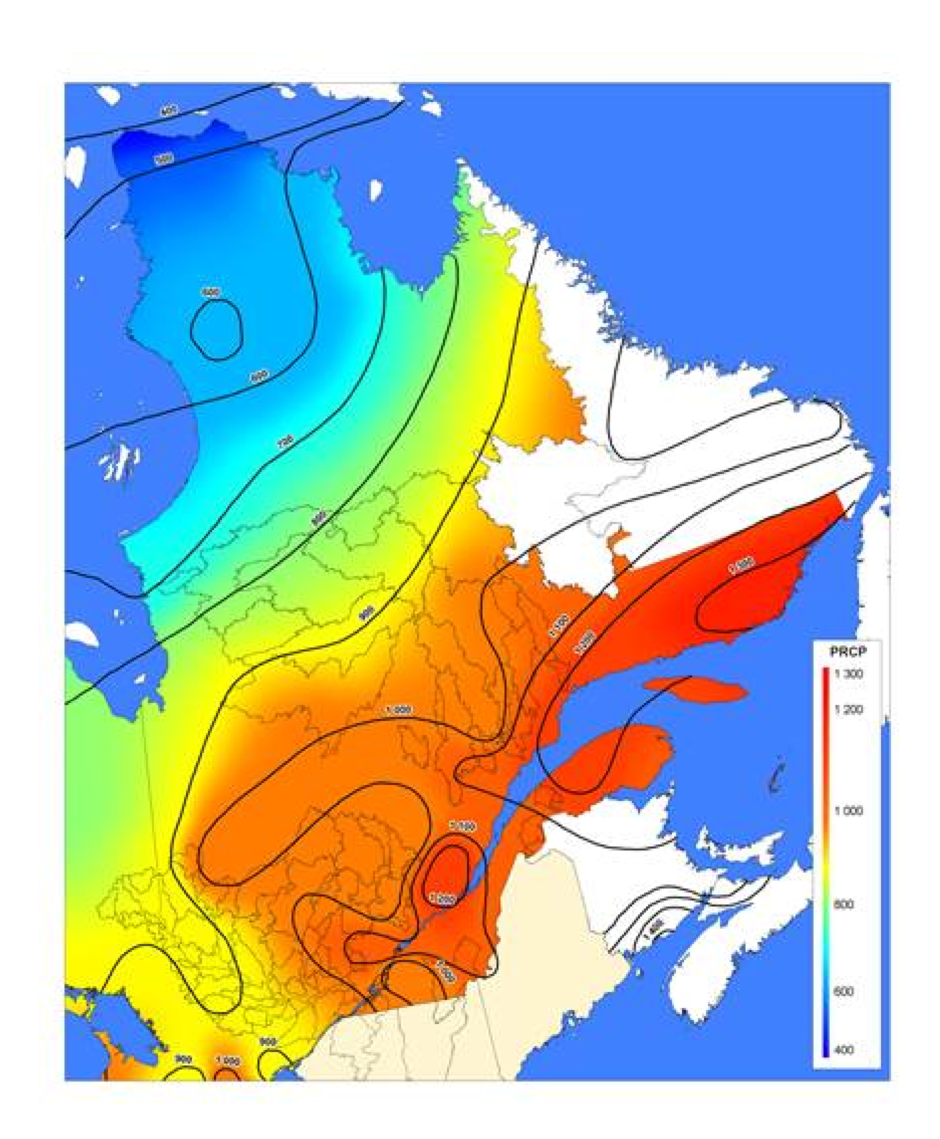
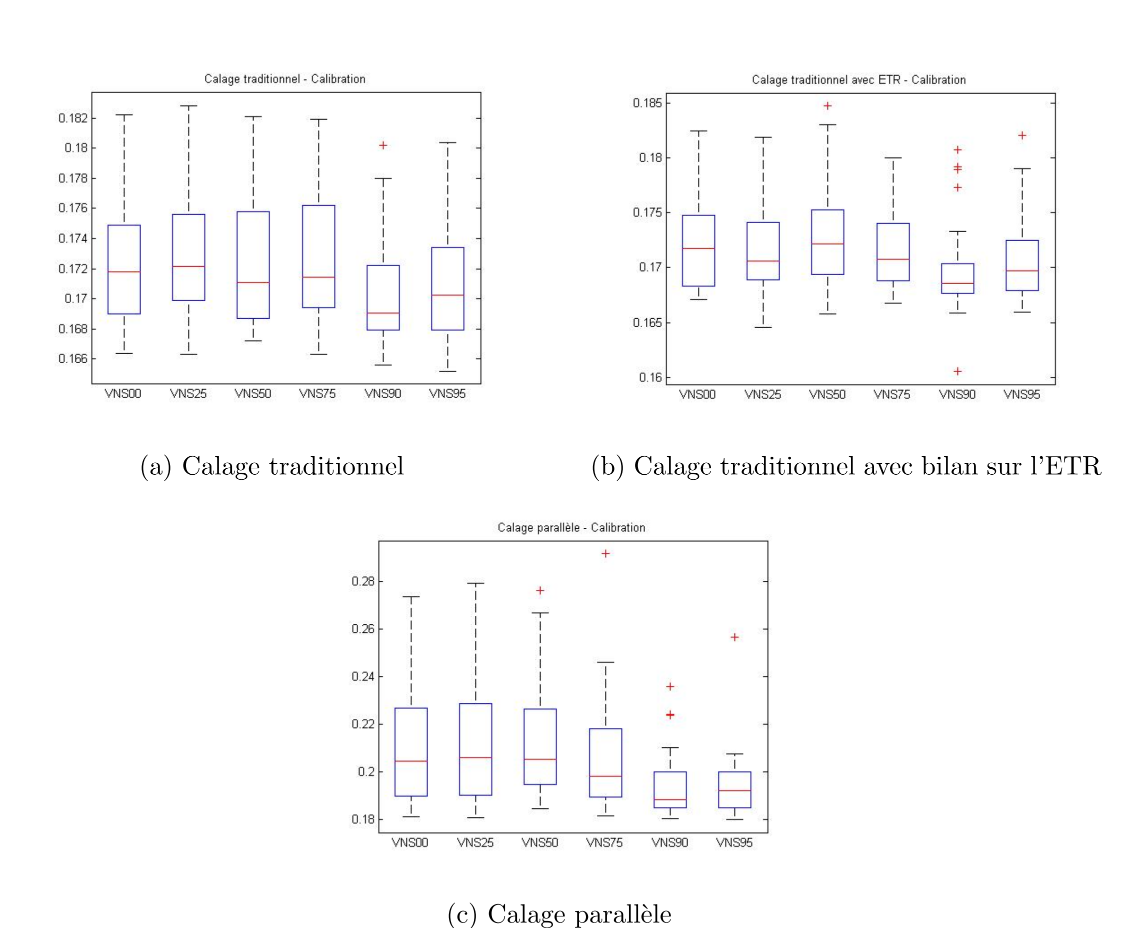


Figure E.2 Précipitations dérivées.

ANNEXE F

CHOIX DU DEGRÉ D'UTILISATION DE LA VNS



(a) Calage traditionnel

(b) Calage traditionnel avec bilan sur l'ETR

(c) Calage parallèle

Figure F.1 Diagrammes en boîtes pour les six stratégies d'utilisation de la VNS.

ANNEXE G

ANALYSES COMPLÉMENTAIRES LIÉS À LA GÉNÉRALISATION DE LA MÉTHODE ET À LA SÉPARATION DES PHASES

G.1 Application à six autres bassins

Afin de de donner plus de poids aux hypothèses tirées des tests effectués sur le bassin versant Caniapiscau, les trois stratégies de calage ont été appliquées à six bassins supplémentaires (Châteauguay, Romaine, Nottaway, Innuksuac, Grand-Mère et Manic-2) possédant des caractéristiques des plus différentes : superficie, latitude, apports naturels ou gérés, types de sols, végétation, topographie, etc. La totalité des résultats obtenus durant cette partie de la recherche figure dans Minville *et al.* (2012), mais les principaux éléments seront repris ici. Une conclusion quant à la généralisation du comportement de l'approche par contrainte sur l'ensemble du Québec pourra être avancée, basée sur l'hypothèse stipulant que les bassins versants choisis, de par leur diversité, représentent bien l'ensemble des conditions présentes au Québec.

G.1.1 Particularités des bassins choisis

La carte suivante (Figure G.1) vient situer sur la province ces nouveaux bassins ainsi que celui ayant servi dans la première phase d'expérimentation (Caniapiscau). L'ensemble des particularités de chaque région ne sera pas détaillé, mais deux informations méritent d'être spécifiées pour pouvoir juger de la performance globale des bassins convenablement. Ainsi, notons que le bassin Innuksuac n'a que sept ans de données météorologiques à son actif. En optimisant le modèle une année sur deux, cela signifie que ce bassin n'a eu droit qu'à quatre ans pour ajuster sa paramétrisation, ce qui est assez peu. Ensuite, la situation géographique du bassin Châteauguay a joué en sa défaveur, d'où sa piètre performance. Chevauchant la frontière américaine, seule la partie canadienne des données météorologiques était disponible dans la base de données de l'IREQ, et ces données n'étaient pas des plus représentatives étant donnée la diversité physiographique du bassin.

Afin d'expliquer les résultats obtenus par la stratégie de calage traditionnel avec bilan sur l'ETR, la Figure G.2 présente les valeurs de l'Atlas pour les six bassins, ainsi que les valeurs obtenues par le calcul $(1 - \frac{Q}{P})$ réalisé à partir des observations. Il ressort de cette figure que la valeur de l'Atlas du Canada pour le bassin Innuksuac est très éloignée de celle obtenue

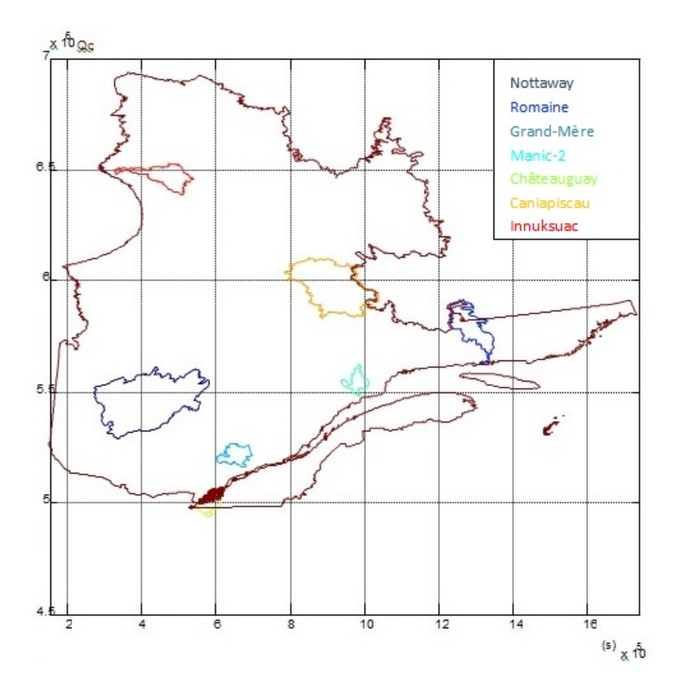


Figure G.1 Carte des sept bassins.

avec les observations. Cela se reflétera probablement dans les résultats de l'optimisation. Il est aussi possible de noter que la moitié des bassins versants présentent une valeur de l'Atlas inférieure aux observations, alors que les trois autres ont une valeur supérieure. L'impact des deux types de comportements pourra donc être étudié.

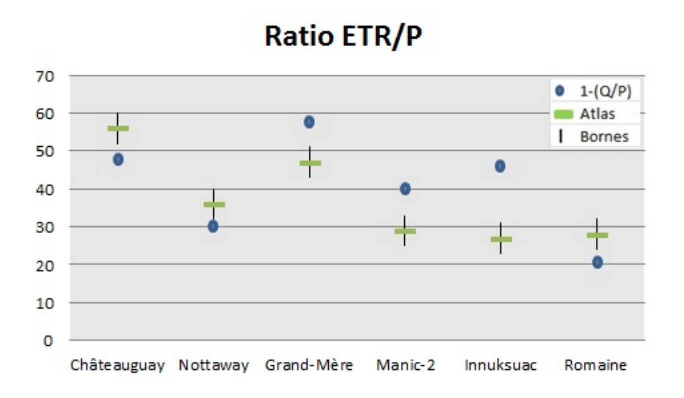


Figure G.2 Graphique des ratios ($\frac{ETR}{P}$) selon l'Atlas encadrées par les bornes utilisées, et de la valeur de ($1 - \frac{Q}{P}$) calculée avec les données météorologiques observées.

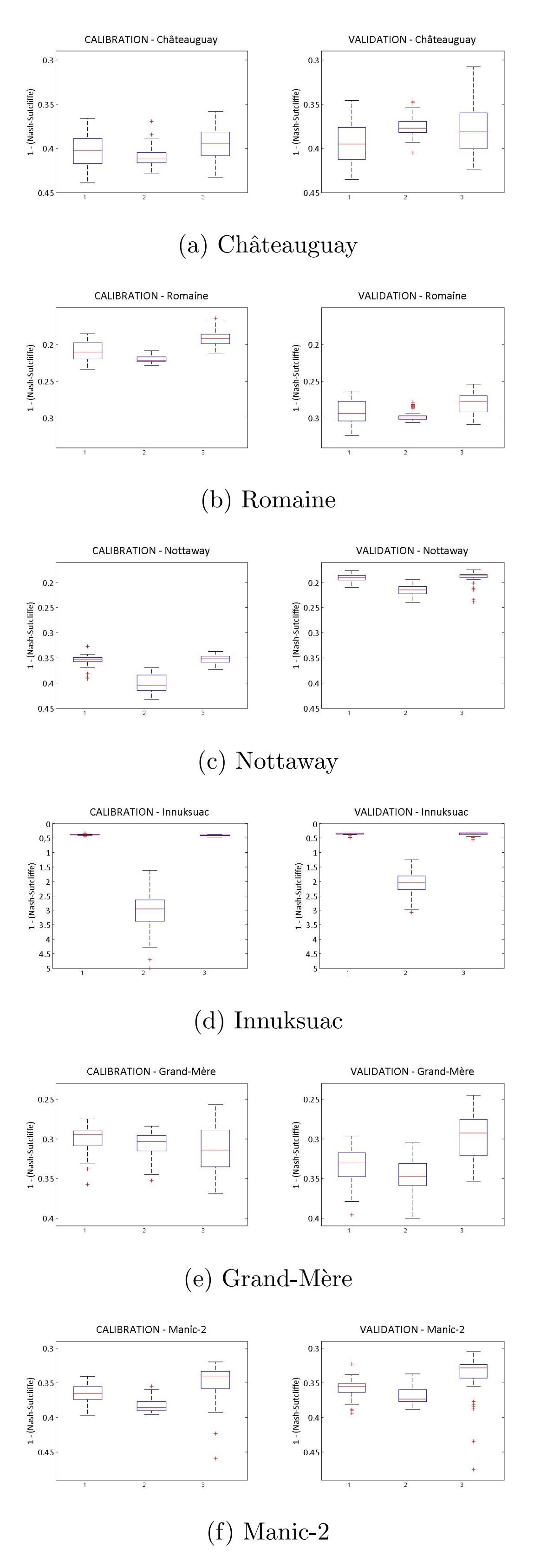
G.1.2 Méthodologie

Avant d'arriver aux résultats concernant la généralisation de la méthode, d'autres analyses de sensibilité ont dû être effectuées, afin d'adapter les contraintes à chacun des bassins. Vu l'hétérogénéité des bassins à l'étude, les analyses de sensibilité ont su se montrer particulièrement utiles. En effet, la réponse aux contraintes étaient très différentes d'un bassin à l'autre, selon la quantité et la qualité des données. Les analyses ont donc permis de personnaliser les contraintes selon la confiance accordée aux données météorologiques et selon la capacité du modèle a reproduire le cycle de l'évapotranspiration. Certains tests ont aussi été mis en œuvre afin de valider le choix du degré d'utilisation de la VNS, qui s'est avéré transposable pour l'ensemble des stratégies. Finalement, la même méthode consistant à comparer cinquante simulations pour chacune des trois stratégies de calage pour chaque bassin a été appliquée. Ainsi, ce sont près de 2 000 optimisations qui ont été lancées pour cette partie du travail, soit l'équivalent de plus d'un an en temps de calcul continu. La parallélisation des tâches ainsi que les possibilités offertes par le centre de calcul CASIR ont cependant permis de réaliser ce travail considérable en une période de temps raisonnable.

G.1.3 Résultats et analyses

Analyse de la robustesse

En premier lieu, il convient de jeter un oeil aux résultats obtenus en calibration et en validation pour chaque des trois stratégies de calage, présentés à la Figure G.3. La principale conclusion est qu'il existe, pour chacun des six bassins, une paramétrisation obtenue par calage parallèle qui supplante, en validation, l'ensemble des solutions, toutes stratégies de calage confondues. Puisqu'un ensemble de paramètres unique est souvent souhaité, cela laisse supposer que le calage parallèle offre une robustesse des solutions surpassant celle des autres méthodes. Une étude de la corrélation entre les solutions obtenues en calibration et en validation a d'ailleurs aussi été menée pour les six bassins. Les graphiques ainsi que les résultats des tests de corrélation apparaissent à la Figure G.4 et aux suivantes. En excluant les bassins Châteauguay (0.46) et Innuksuac (0.67), la corrélation des solutions obtenues suite à l'utilisation du calage parallèle varie entre 0.95 et 0.99. Les piètres résultats des deux bassins précédemment nommés pourraient s'expliquer par les raisons mentionnées plus haut, soit le manque de données et leur qualité critiquable.



(a) Châteauguay

(b) Romaine

(c) Nottaway

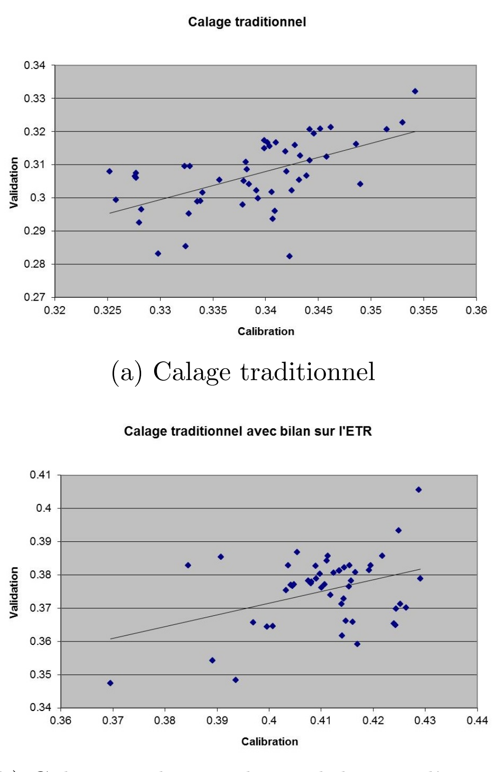
(d) Innuksuac

(e) Grand-Mère

(f) Manic-2

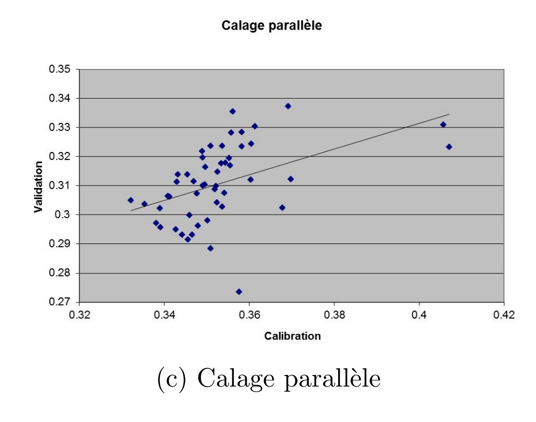
Figure G.3 Résultats en calibration et en validation pour les six bassins.

Pour le bassin versant Châteauguay, les valeurs du coefficient de corrélation sont approximativement de 0.575, 0.388 et 0.458, respectivement, pour les stratégies de calage traditionnel, de calage traditionnel avec bilan sur l'ETR et de calage parallèle.



(a) Calage traditionnel

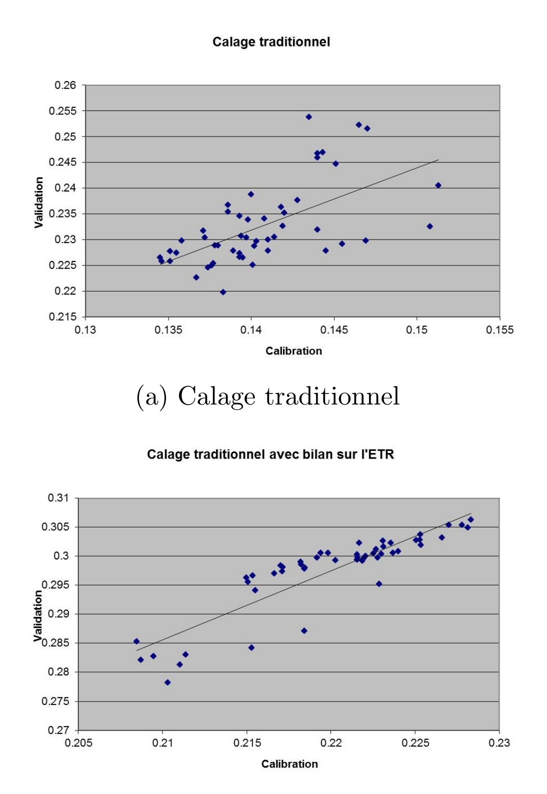
(b) Calage traditionnel avec bilan sur l'ETR



(c) Calage parallèle

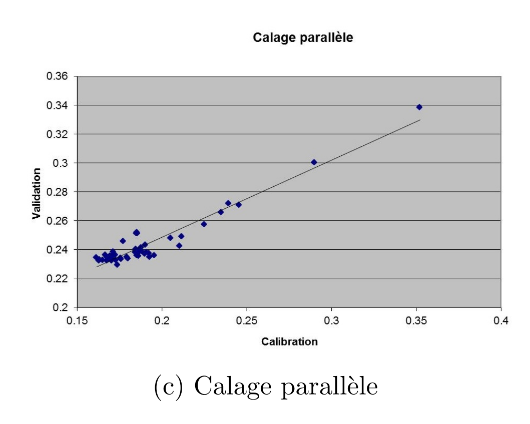
Figure G.4 Corrélation dans les solutions pour le bassin Châteauguay.

Pour le bassin versant Romaine, les valeurs du coefficient de corrélation sont approximativement de 0.602, 0.885 et 0.959, respectivement, pour les stratégies de calage traditionnel, de calage traditionnel avec bilan sur l'ETR et de calage parallèle.



(a) Calage traditionnel

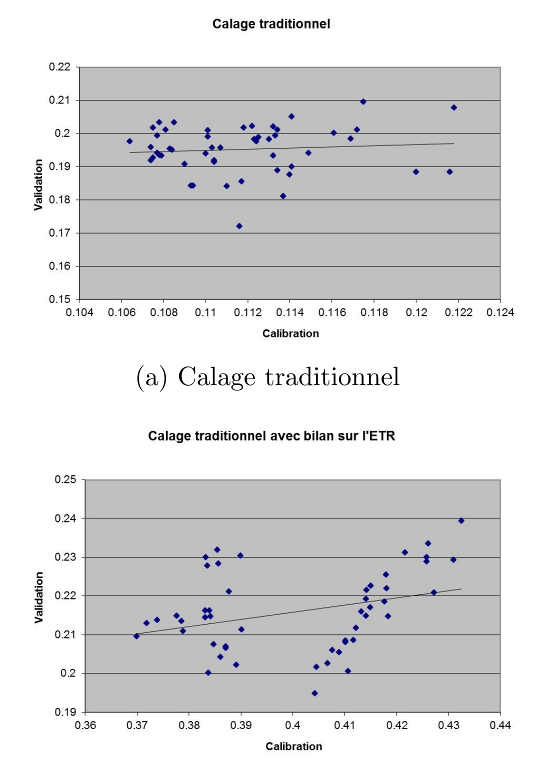
(b) Calage traditionnel avec bilan sur l'ETR



(c) Calage parallèle

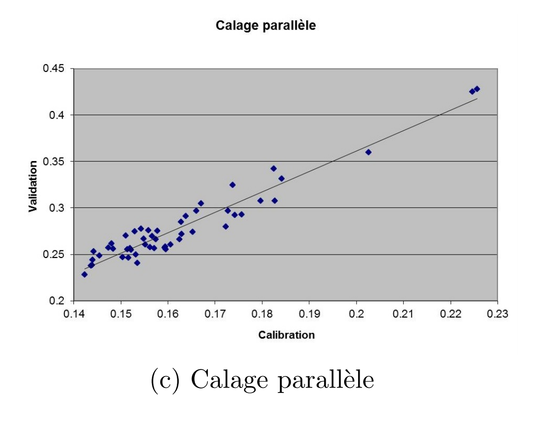
Figure G.5 Corrélation dans les solutions pour le bassin Romaine.

Pour le bassin versant Nottaway, les valeurs du coefficient de corrélation sont approximativement de 0.091, 0.322 et 0.960, respectivement, pour les stratégies de calage traditionnel, de calage traditionnel avec bilan sur l'ETR et de calage parallèle.



(a) Calage traditionnel

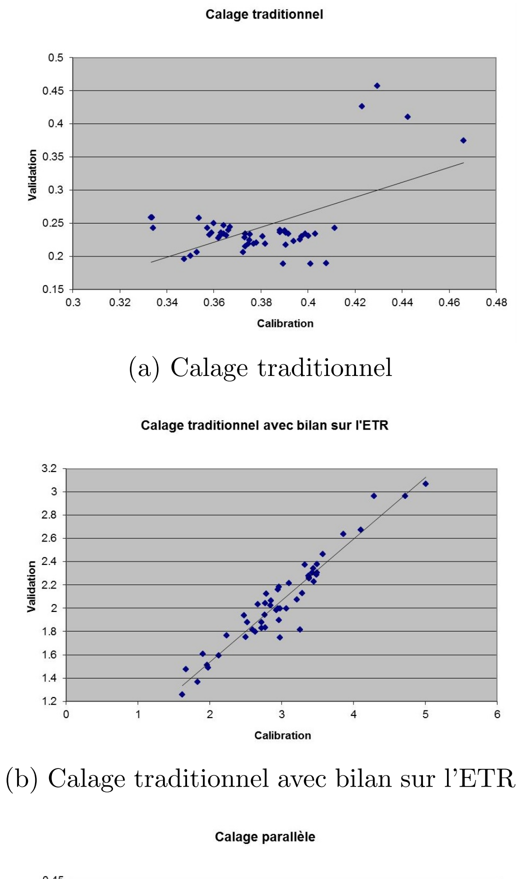
(b) Calage traditionnel avec bilan sur l'ETR



(c) Calage parallèle

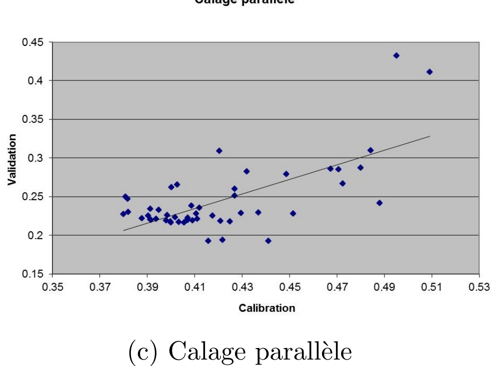
Figure G.6 Corrélation dans les solutions pour le bassin Nottaway.

Pour le bassin versant Innuksuac, les valeurs du coefficient de corrélation sont approximativement de 0.549, 0.957 et 0.675, respectivement, pour les stratégies de calage traditionnel, de calage traditionnel avec bilan sur l'ETR et de calage parallèle.



(a) Calage traditionnel

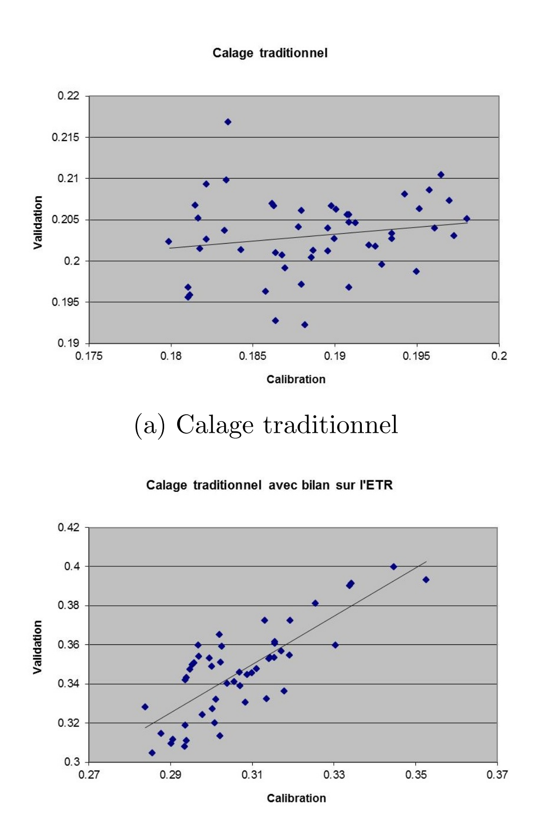
(b) Calage traditionnel avec bilan sur l'ETR



(c) Calage parallèle

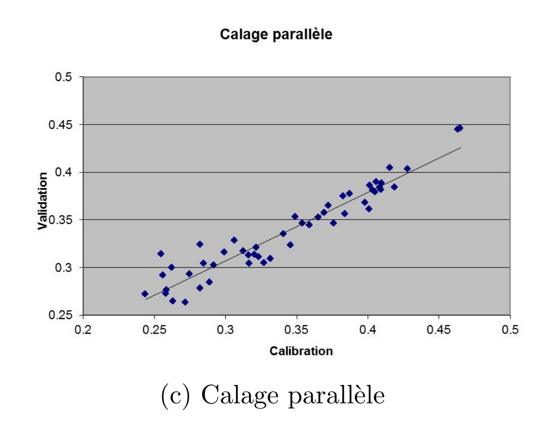
Figure G.7 Corrélation dans les solutions pour le bassin Innuksuac.

Pour le bassin versant Grand-Mère, les valeurs du coefficient de corrélation sont approximativement de 0.179, 0.791 et 0.955, respectivement, pour les stratégies de calage traditionnel, de calage traditionnel avec bilan sur l'ETR et de calage parallèle.



(a) Calage traditionnel

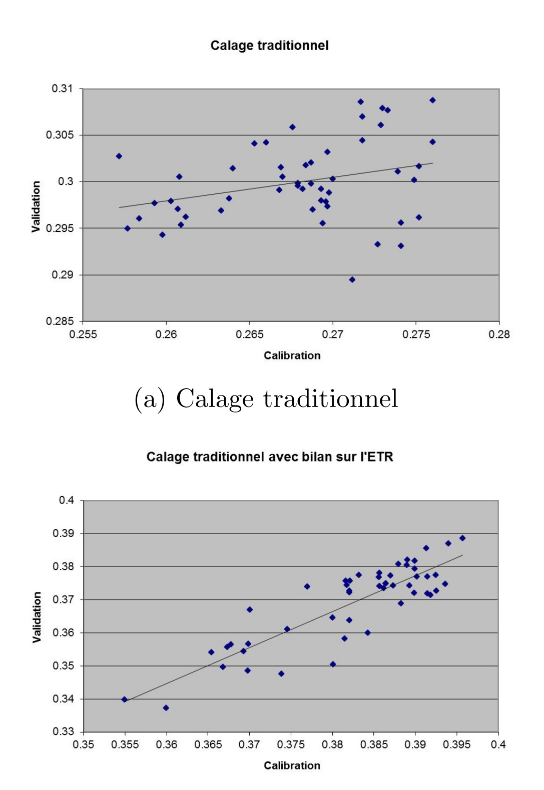
(b) Calage traditionnel avec bilan sur l'ETR



(c) Calage parallèle

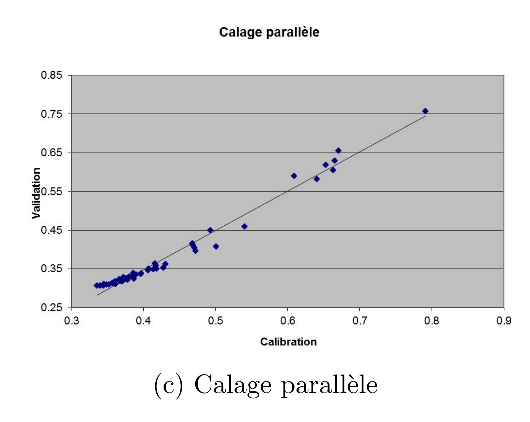
Figure G.8 Corrélation dans les solutions pour le bassin Grand-Mère.

Pour le bassin versant Manic-2, les valeurs du coefficient de corrélation sont approximativement de 0.308, 0.864 et 0.991, respectivement, pour les stratégies de calage traditionnel, de calage traditionnel avec bilan sur l'ETR et de calage parallèle.



(a) Calage traditionnel

(b) Calage traditionnel avec bilan sur l'ETR



(c) Calage parallèle

Figure G.9 Corrélation dans les solutions pour le bassin Manic-2.

Quant à la stratégie de calage traditionnel avec ETR, les diagrammes en boîtes montrent généralement une détérioration en termes de critère de Nash-Sutcliffe modifié $(1 - ns)$. Tel que prévu d'ailleurs, le bassin Innuksuac a subit la plus forte dégradation et l'hypothèse selon laquelle un écart élevé entre la valeur de l'Atlas et la valeur du ratio $1 - \frac{Q}{P}$ entraîne une diminution plus marquée de la fonction objectif s'en voit renforcée. L'écart positif ou négatif entre ces deux valeurs ne semble toutefois pas avoir d'impact.

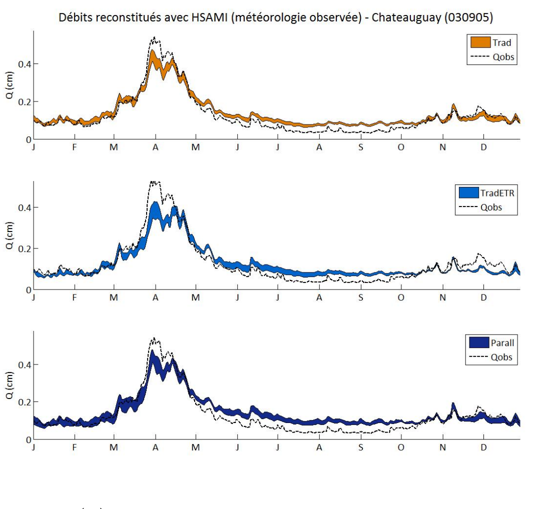
Analyse des hydrogrammes

Les hydrogrammes de tous les bassins, pour chacune des stratégies de calage sont présentés aux Figures G.10 et G.11. De ces figures, trois constatations doivent être faites. En premier lieu, les hydrogrammes des bassins versants Châteauguay et Innuksuak sont une autre preuve d'un problème au niveau des données. Avec toutes les stratégies de calage, et particulièrement avec le calage traditionnel avec bilan sur l'ETR, les débits sont systématiquement mal modélisés. Ensuite, de façon générale, ces figures montrent une similarité entre les hydrogrammes produits par les stratégies de calage traditionnel et traditionnel avec ETR. La stratégie de calage parallèle, quant à elle, montre une plus grande disparité entre les solutions au niveau de la modélisation des apports. Finalement, alors qu'un retard dans la montée de crue était observé pour le bassin Caniaspicau, les stratégies de calage traditionnel et traditionnel avec bilan sur l'ETR semblent devancer la crue pour plusieurs bassins. Cette anticipation de la crue semble moins prononcée pour le calage parallèle.

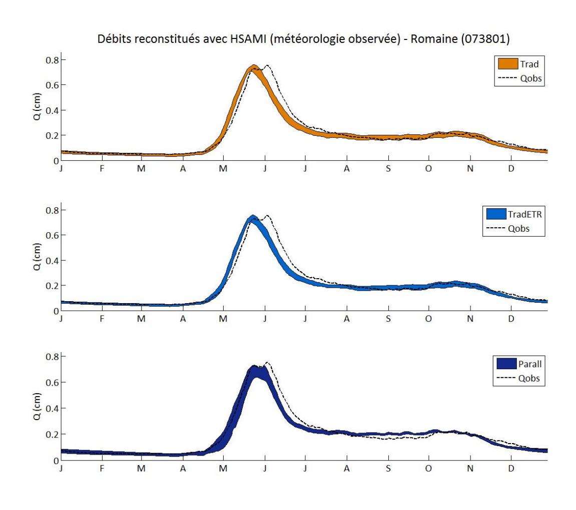
Analyse des évapogrammes

Afin de porter des conclusions sur la modélisation de l'évapotranspiration, les cycles de l'évapotranspiration réelle produits avec la météorologie de la simulation climatique *aft* sont présentés aux Figures G.12 et G.13, où les courbes des trois stratégies de calage sont superposées. Les évapogrammes présentant chacune des stratégies individuellement se trouvent à l'Annexe H suivante, tandis que les évapogrammes produits avec la météorologie observée ne seront pas présentés. Ceux-ci ont été générés, mais ils ne donnaient pas plus d'information que ceux inclus dans ce chapitre. Ils sont cependant présents en annexe de Minville *et al.* (2012).

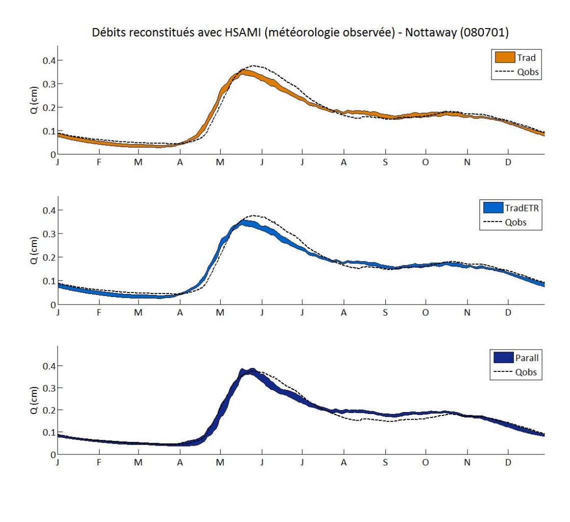
Deux types d'informations seront tirées de ces figures. Premièrement, la largeur de l'enveloppe des solutions renseigne sur la variabilité de chaque stratégie de calage. Ensuite, une



(a) Bassin versant Châteauguay

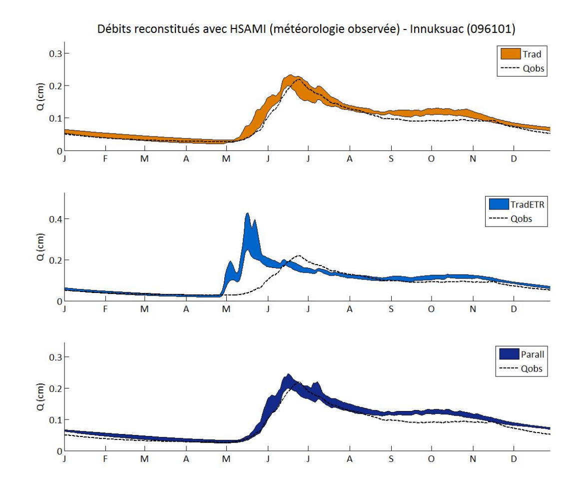


(b) Bassin versant Romaine

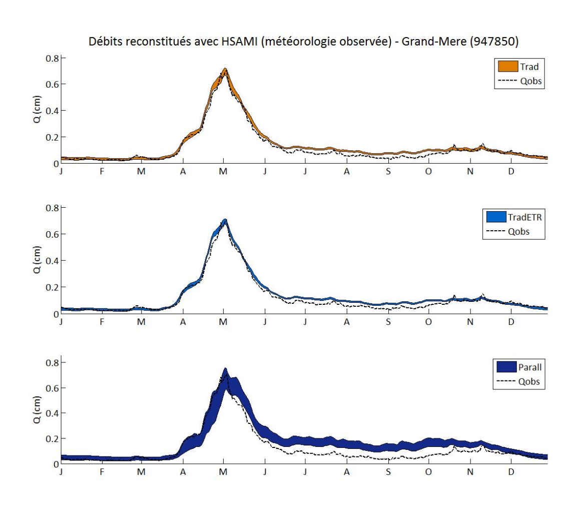


(c) Bassin versant Nottaway

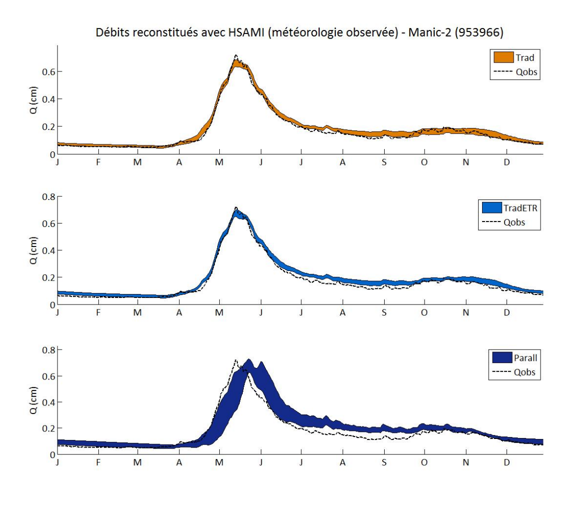
Figure G.10 Hydrogrammes des bassins versants Châteauguay, Romaine et Nottaway.



(a) Bassin versant Innuksuac

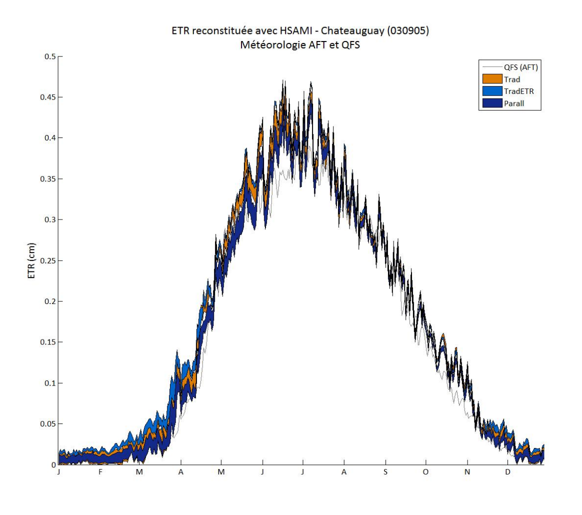


(b) Bassin versant Grand-Mère

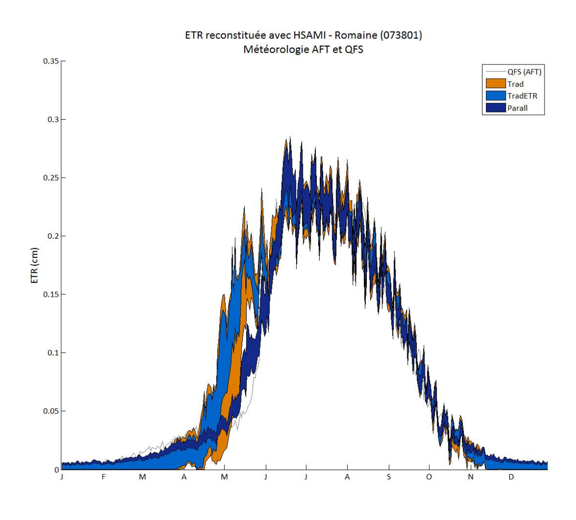


(c) Bassin versant Manic-2

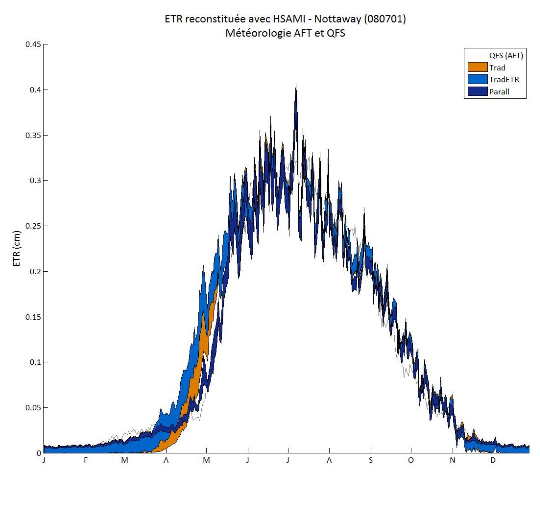
Figure G.11 Hydrogrammes des bassins versants Innuksuac, Grand-Mère et Manic-2.



(a) Bassin versant Châteauguay

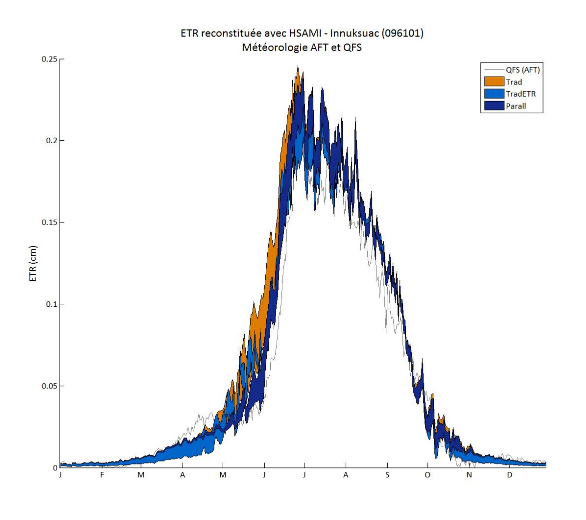


(b) Bassin versant Romaine

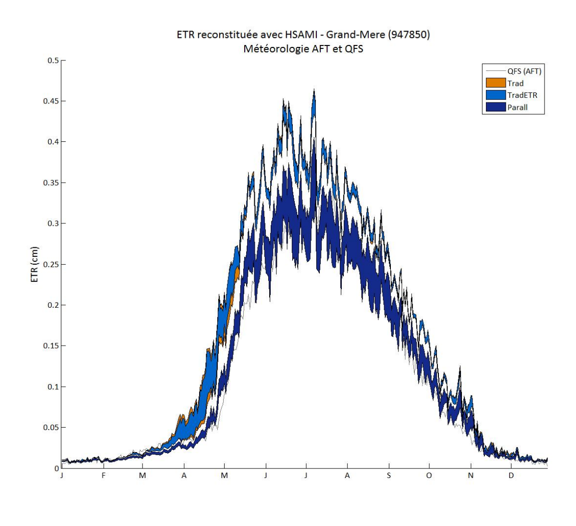


(c) Bassin versant Nottaway

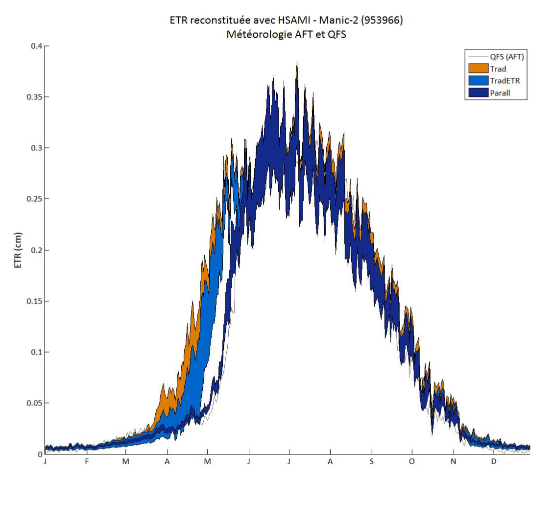
Figure G.12 Évapogrammes des bassins versants Châteauguay, Romaine et Nottaway.



(a) Bassin versant Innuksuac



(b) Bassin versant Grand-Mère



(c) Bassin versant Manic-2

Figure G.13 Évapogrammes des bassins versants Innuksuac, Grand-Mère et Manic-2.

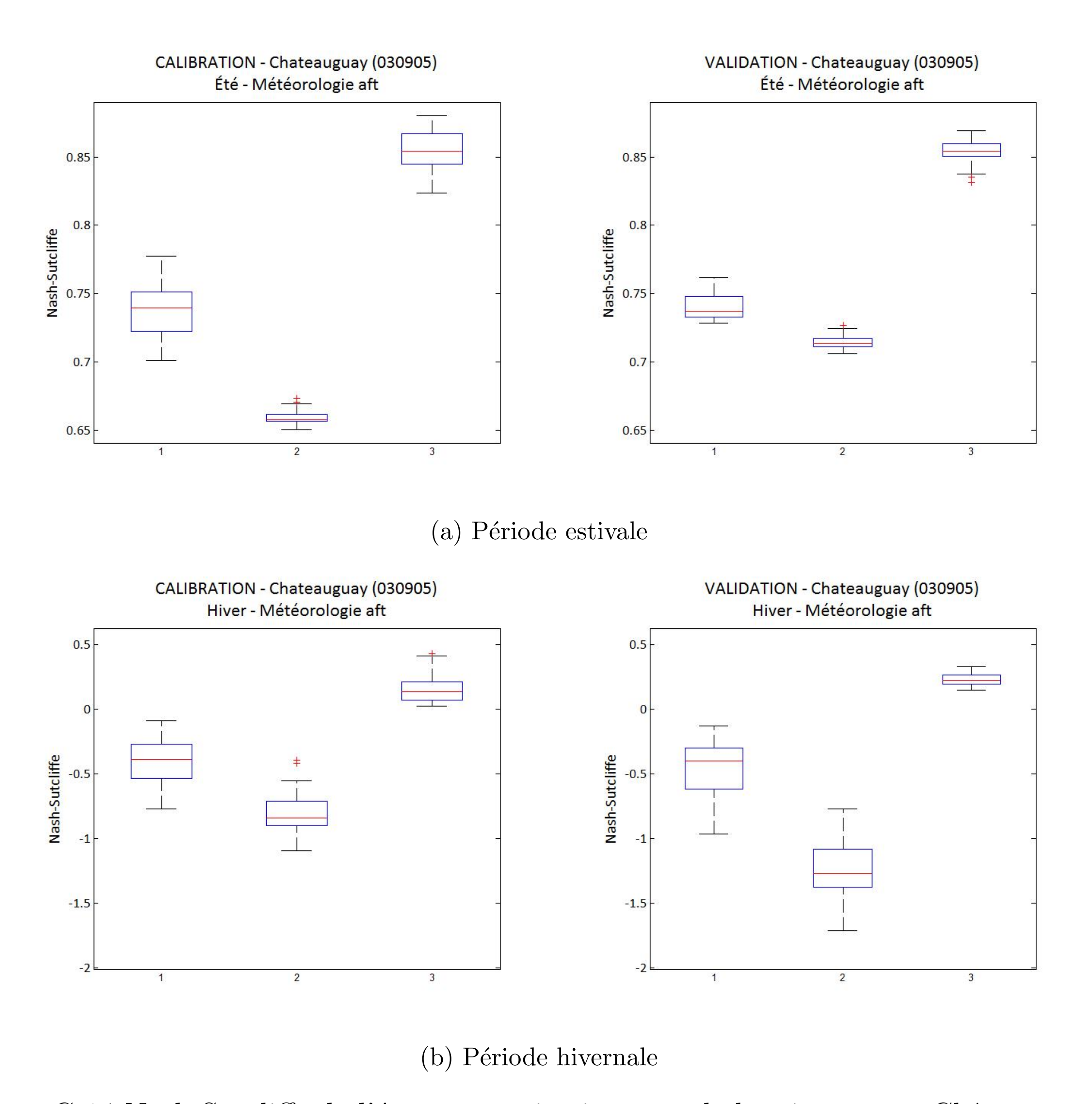
attention particulière sera portée à la représentation juste de la variable Qfs. Au niveau de la dispersion des solutions, le calage parallèle offre des simulations plus près les unes des autres pour la période hivernale. Pour la période estivale cependant, les trois stratégies sont généralement comparables en terme de variabilité, sauf pour les bassins Grand-Mère et Manic-2, où le calage parallèle est un peu plus variable. En regardant la représentation du cycle de la variable Qfs cependant, le calage parallèle supplante toutes les autres stratégies, surtout en période hivernale où le cycle est déphasé de plusieurs semaines par les stratégies de calage traditionnel et traditionnel avec bilan sur l'ETR.

Respect des contraintes

Les diagrammes en boîtes présentant les valeurs du ratio $\frac{ETR}{P}$ ne seront pas intégrés dans ce rapport. Ces figures ne contribuent pas au choix d'une stratégie puisque les deux sources de données valorisées dans cette recherche, soit l'Atlas du Canada et le modèle climatique MRCC, véhiculent généralement de l'information contradictoire quant au bilan d'évapotranspiration. Ne sachant pas laquelle est la plus près de notre réalité, il est impossible de statuer. Cela permet tout de même de proscrire l'utilisation des deux méthodes de calage, traditionnel avec bilan sur l'ETR et parallèle, pour contraindre une même optimisation vu leur conflictualité.

Pour juger de la forme du cycle, les diagrammes en boîtes présentant la valeur du critère de Nash-Sutcliffe pour les périodes hivernale et estivale sont présentés à la Figure G.14 et aux suivantes[1]. Plusieurs constatations peuvent être faites à partir de ces figures. Premièrement, tout comme pour Caniapiscau, les valeurs obtenues en période estivale sont toujours meilleures que celles associées à la période hivernale, qui sont souvent négatives. Sauf pour le bassin versant Innuksuac, la stratégie de calage parallèle offre systématiquement des solutions aux valeurs supérieures pour cette partie de l'année. Ces valeurs sont d'ailleurs toujours positives, signe que les simulations obtenues par le calage parallèle offrent une meilleure performance que l'utilisation de la moyenne, contrairement aux deux autres stratégies. Pour le bassin Innuksuac, la stratégie de calage traditionnel avec bilan sur l'ETR offre généralement de meilleurs résultats que les deux autres méthodes, sauf en validation sur la période estivale où le calage parallèle obtient quelques solutions surpassant les autres.

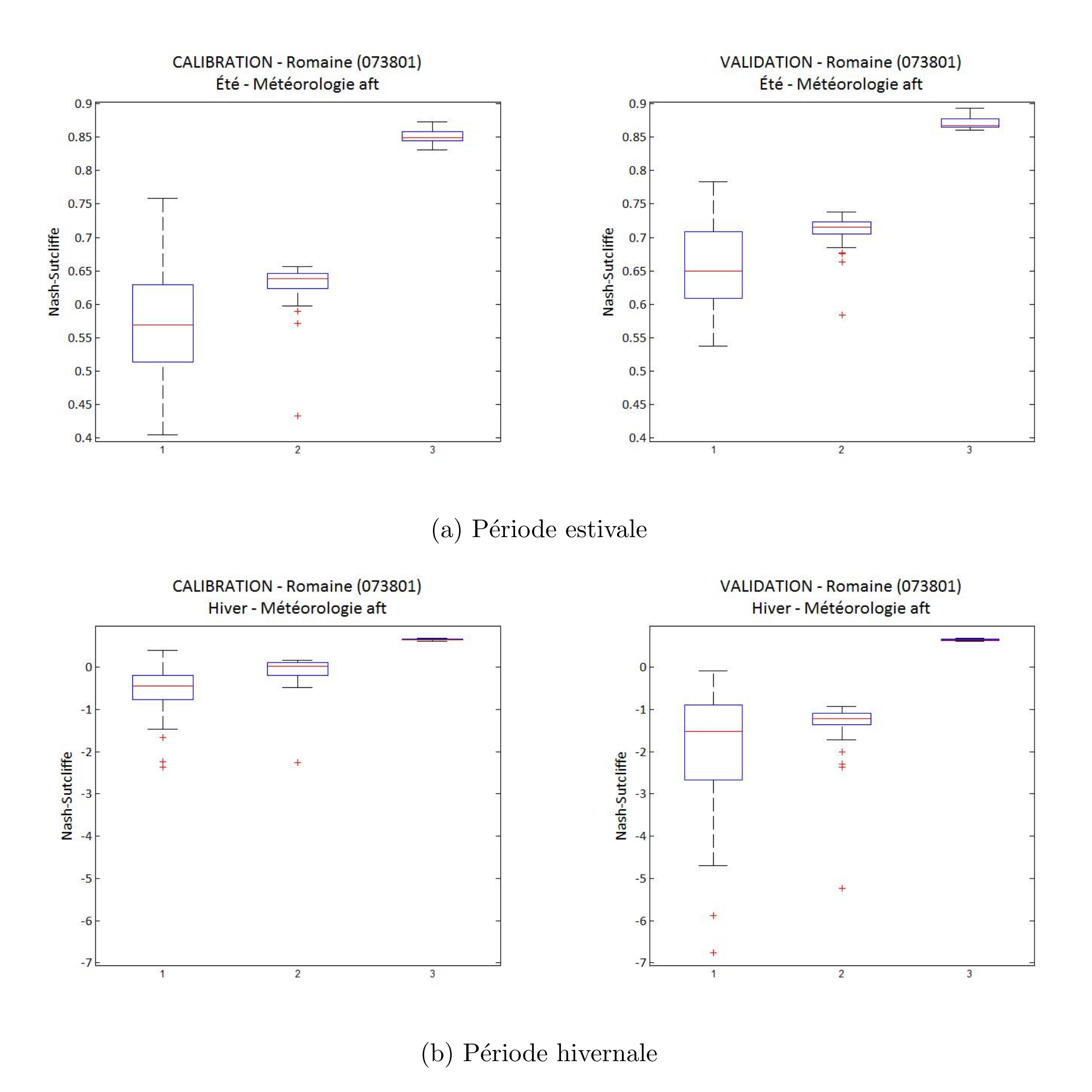
1. Il est important de noter que les figures de cette annexe présentent des valeurs du critère de Nash-Sutcliffe, et non des $1 - ns$. Une valeur près de 1 est donc signe d'une bonne représentation du cycle de l'évapotranspiration, au contraire d'une valeur près de 0. Une valeur négative est signe qu'une utilisation de la simple moyenne serait plus judicieuse que la série ayant été simulée.



(a) Période estivale

(b) Période hivernale

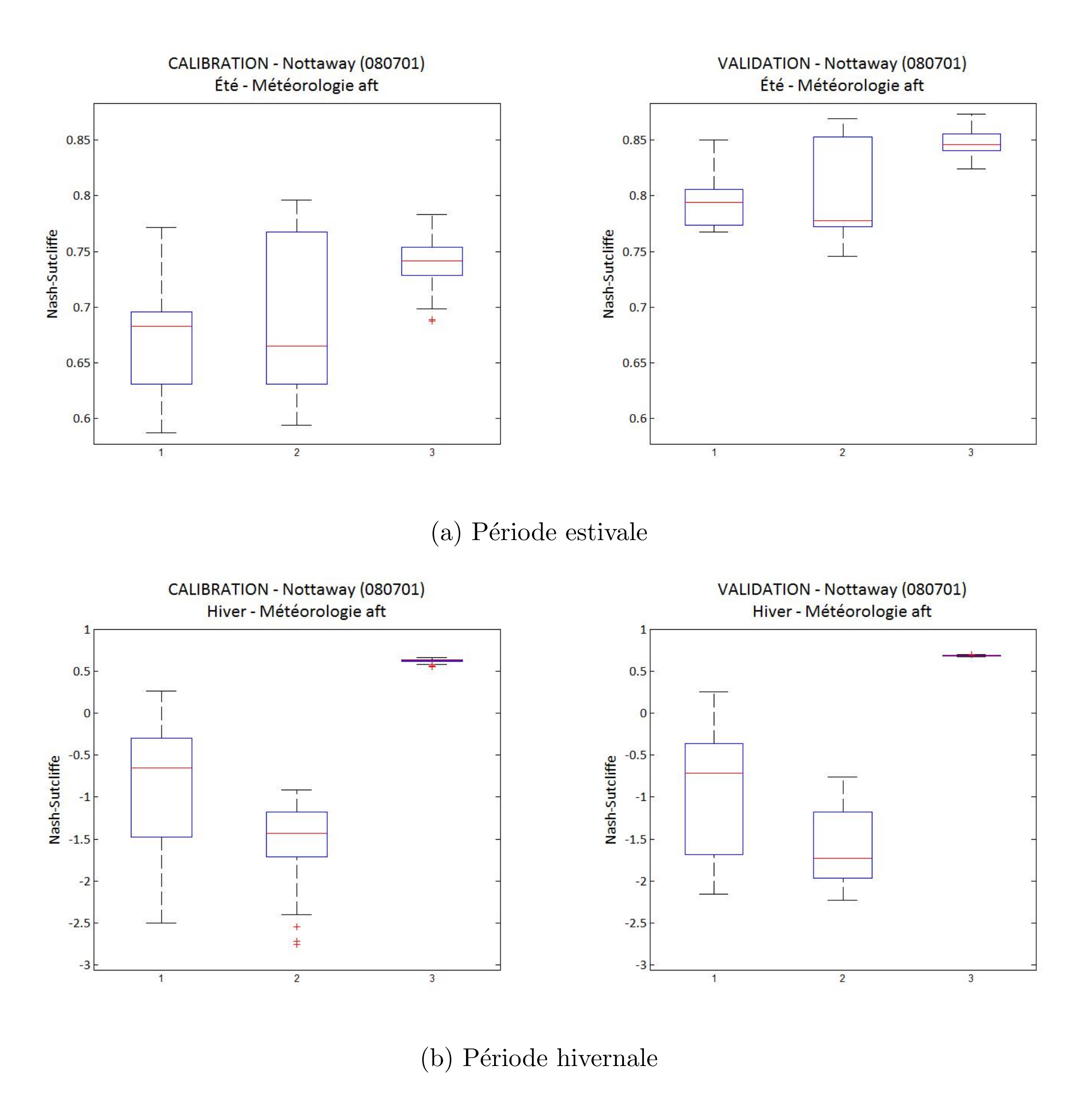
Figure G.14 Nash-Sutcliffe de l'évapotranspiration pour le bassin versant Châteauguay.



(a) Période estivale

(b) Période hivernale

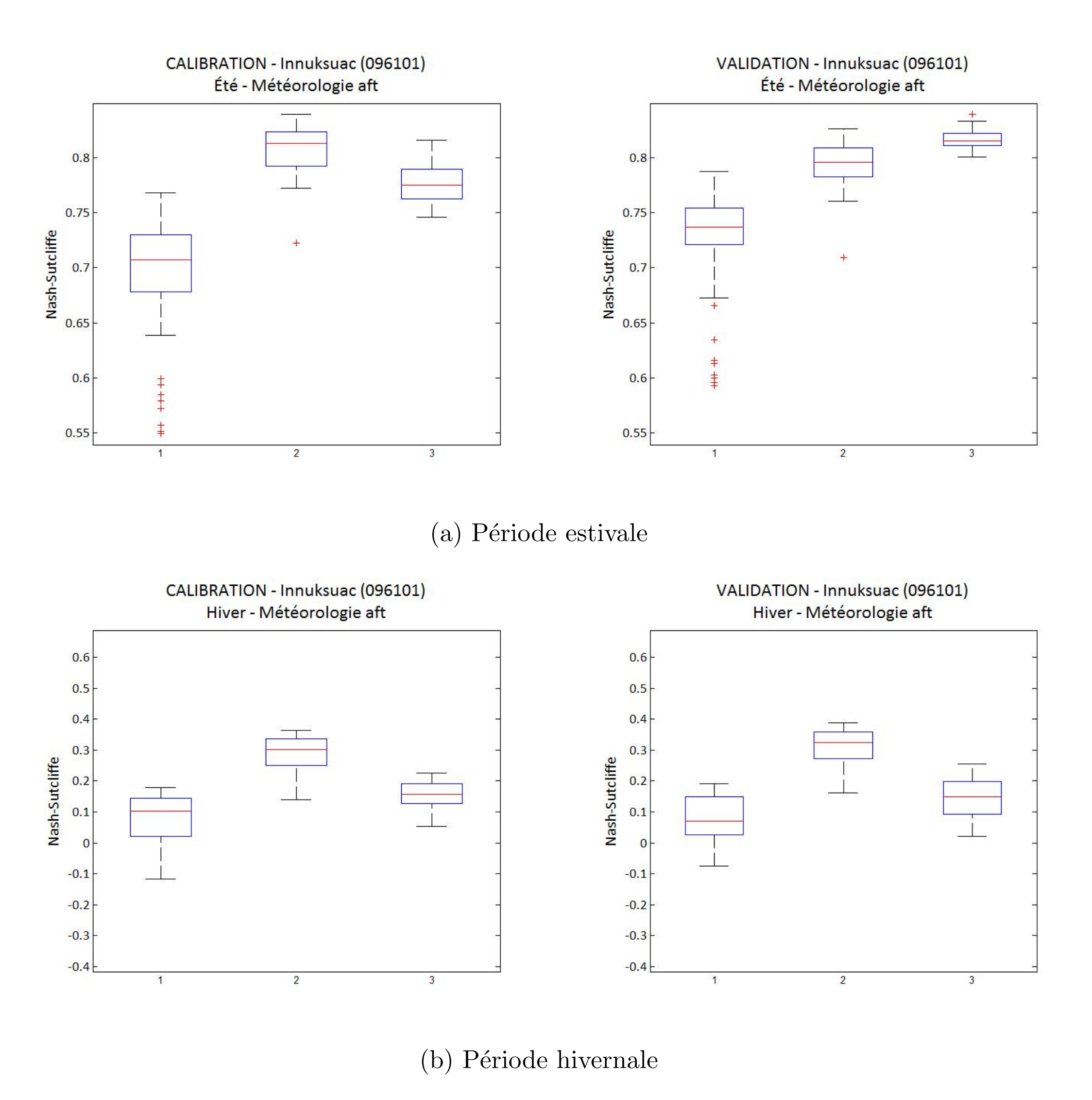
Figure G.15 Nash-Sutcliffe de l'évapotranspiration pour le bassin versant Romaine.



(a) Période estivale

(b) Période hivernale

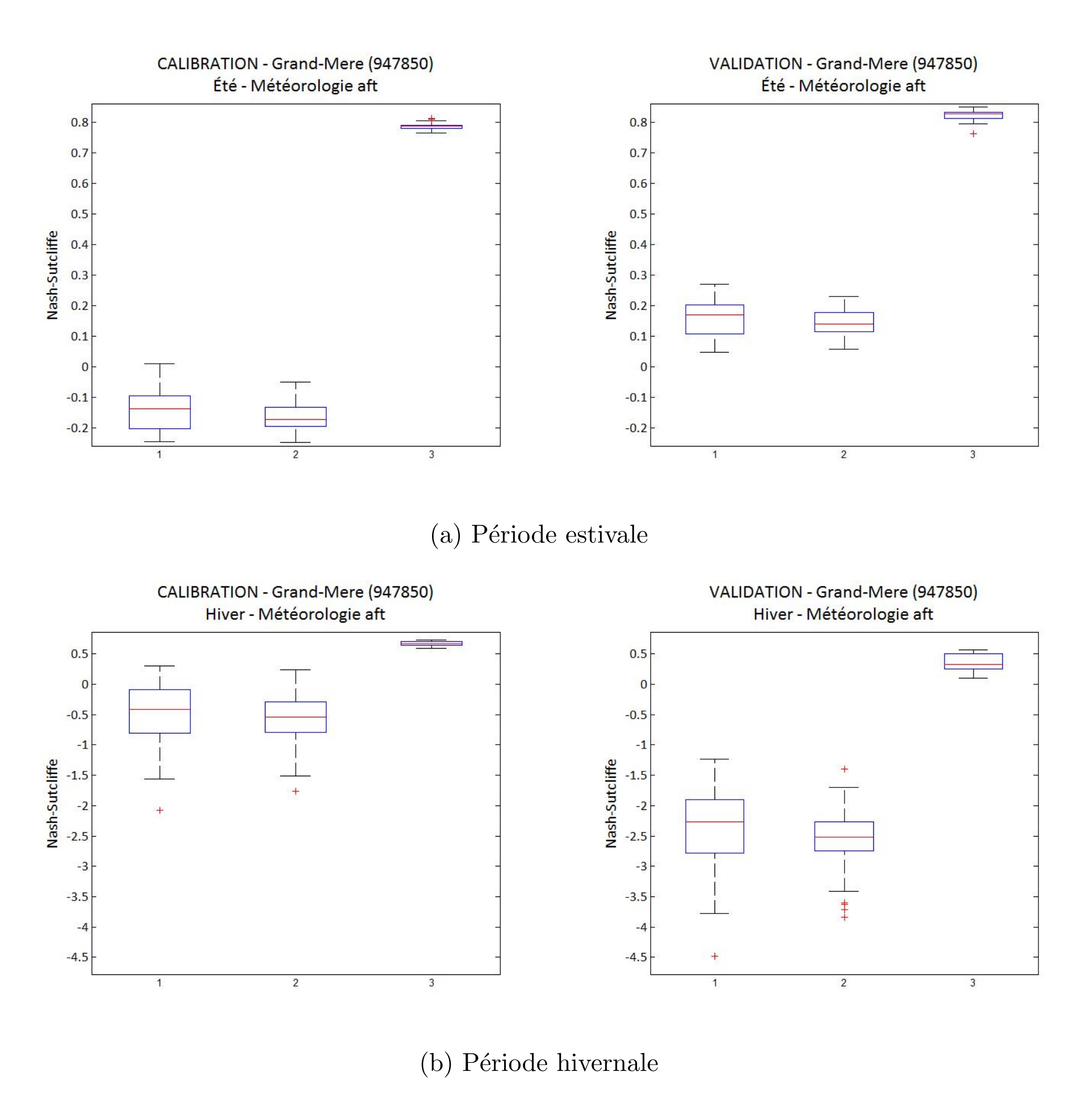
Figure G.16 Nash-Sutcliffe de l'évapotranspiration pour le bassin versant Nottaway.



(a) Période estivale

(b) Période hivernale

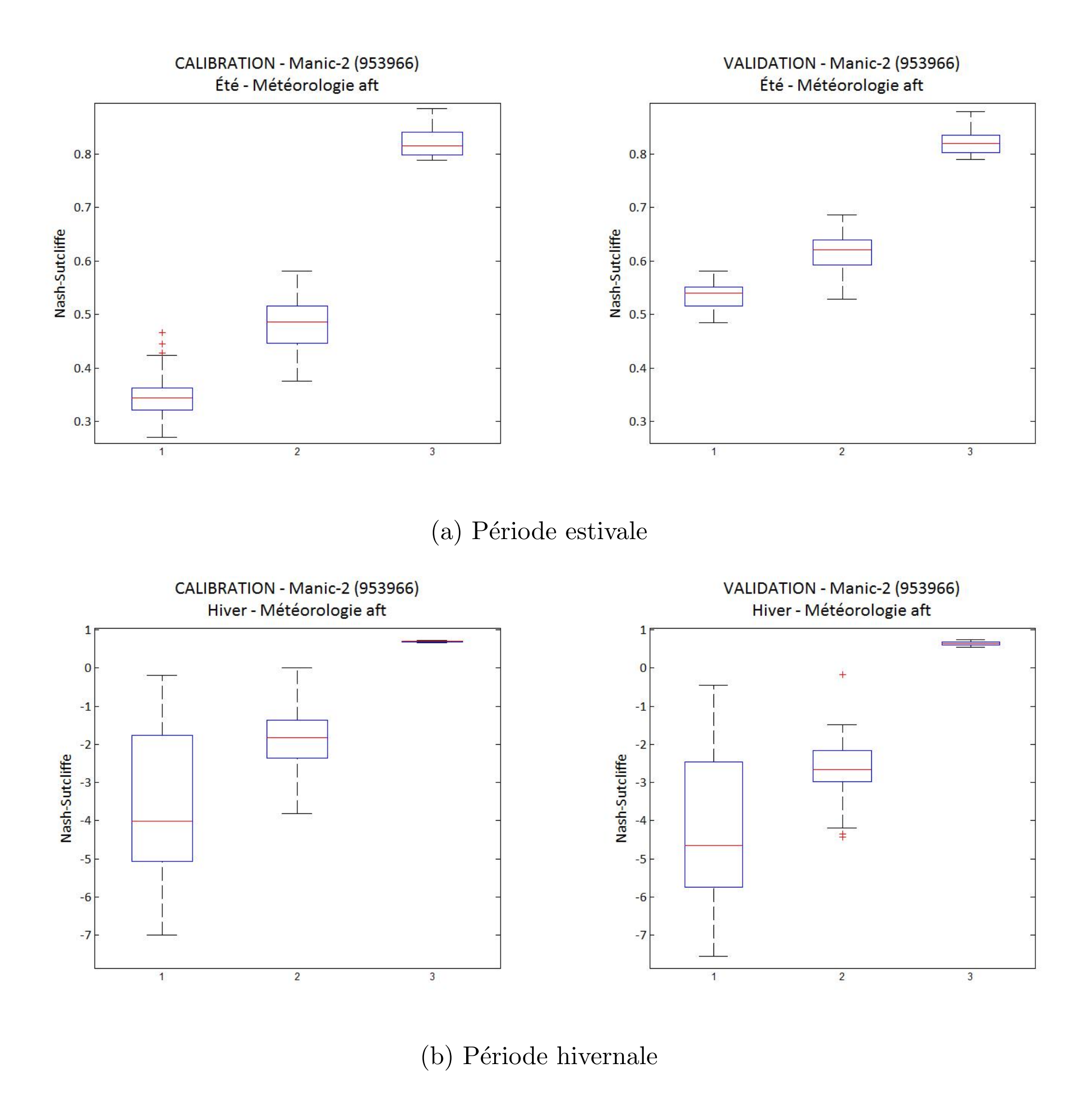
Figure G.17 Nash-Sutcliffe de l'évapotranspiration pour le bassin versant Innuksuac.



(a) Période estivale

(b) Période hivernale

Figure G.18 Nash-Sutcliffe de l'évapotranspiration pour le bassin versant Grand-Mère.



(a) Période estivale

(b) Période hivernale

Figure G.19 Nash-Sutcliffe de l'évapotranspiration pour le bassin versant Manic-2.

G.1.4 Conclusions et généralisation de la méthode

En combinant les résultats présentés dans cette section et ceux du chapitre 5 obtenus pour le bassin versant Caniapiscau, il est possible de dresser un portrait global des différentes méthodes de calage et de statuer sur la possible généralisation de la méthode à l'ensemble du Québec.

Ainsi, la stratégie de calage parallèle a fourni la meilleure solution en période de validation pour six des sept bassins. Dans l'optique où un seul jeu de paramètres doit être considéré pour une application en changements climatiques, et puisqu'une forte corrélation entre calibration et validation a été observée pour cinq des bassins, il est possible de conclure que la stratégie de calage parallèle offre une paramétrisation des plus robustes. Cette stratégie a aussi permis une représentation de la variable Qfs plus fidèle que celle produite par les deux autres méthodes. Les solutions trouvées par cette stratégie de calage proposent donc une paramétrisation du modèle permettant une représentation du processus d'évapotranspiration orientée vers la physique sans compromettre la projection des apports naturels. Un bémol doit cependant être apporté quant à la synchronisation de la montée de crue, qui est parfois retardée par le calage parallèle. Une explication a tout de même été fournie impliquant la séparation des phases pluie et neige.

Aucun avantage n'a pu être associé avec la méthode de calage traditionnel. Non seulement celle-ci ne propose-t-elle pas une meilleure adéquation entre débits simulés et observés que les autres stratégie, mais elle s'est aussi révélée peu apte à reproduire les quantités globales d'évapotranspiration telles que définies par le bilan sur l'ETR et l'allure générale du cycle produit par un modèle climatique. Dans un contexte de projections en changements climatiques, cette stratégie devrait donc être écartée de facto.

Finalement, la méthode de calage traditionnel avec bilan sur l'ETR pourrait être considérée comme un entre-deux. Elle permet de recentrer le bilan de l'évapotranspiration réelle sur les précipitations lorsque la qualité des débits est incertaine. Cependant, cela se fait parfois au détriment de la reproduction des débits, surtout pour les bassins dont la valeur de l'Atlas du Canada est très éloignée de celle dérivée des observations.

De ces conclusions, la décision d'utiliser la stratégie de calage parallèle pour une étude des impacts des changements climatiques a été prise. Les paramétrisations obtenues par cette stratégie semblent robustes lorsqu'elles sont confrontées à des données n'ayant pas servi à la calibration, ce qui laisse présager un comportement fiable en contexte de changements

climatiques. La relative stabilité des résultats entre les bassins versants présentant des caractéristiques des plus variables permet de conclure à une possible généralisation de l'application de la méthode à l'ensemble du Québec.

G.2 Modification de la séparation des phases

La modélisation de la neige s'étant révélée différente pour les modèles climatique et hydrologique, et cela ayant des incidences importantes notamment sur la modélisation du processus de l'évapotranspiration, ce problème a été traité de façon prioritaire. Ainsi, une première vague de tests a été réalisée en parallèle à ce projet et les résultats préliminaires sont présentés dans cette section. Rappelons qu'après avoir analysé les évapogrammes obtenus suite au calage du modèle HSAMI, force était de constater un retard du modèle hydrologique au niveau de l'évapotranspiration au début de l'hiver (l'exemple le plus flagrant est présenté à la Figure G.20). Après investigation, une divergence entre la quantité de neige simulée par le modèle hydrologique et par le modèle climatique fut constatée, comme le montre la Figure G.21. Il a été démontré que la quantité de précipitations totales des deux modèles était comparable, mais que la séparation des phases pluie-neige, modélisée différemment pour les deux types de modèles, posait problème.

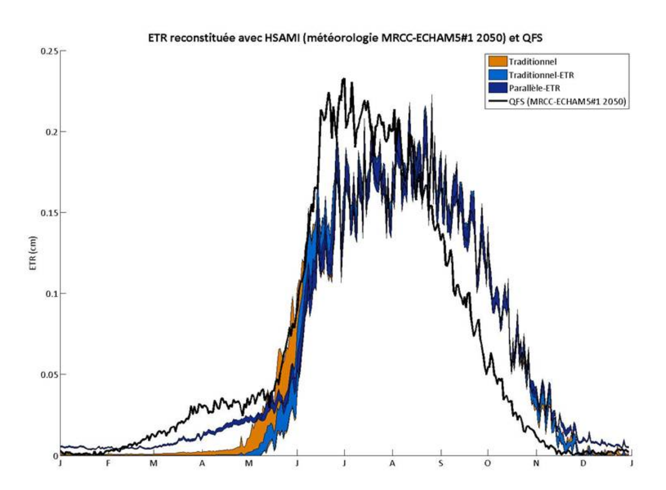


Figure G.20 Évapogramme simulé avec la météo de la simulation ECHAM5#1.

La différence entre les deux méthodes employées est présentée aux Figures G.22 et G.23, qui proposent respectivement un cycle annuel moyen ainsi qu'une moyenne annuelle sur 140 ans. Pour le modèle HSAMI, la séparation des précipitations est effectuée par la méthode dite [-2 ; 2]. Lorsque la température moyenne journalière est inférieure à -2° C, l'ensemble des pré-

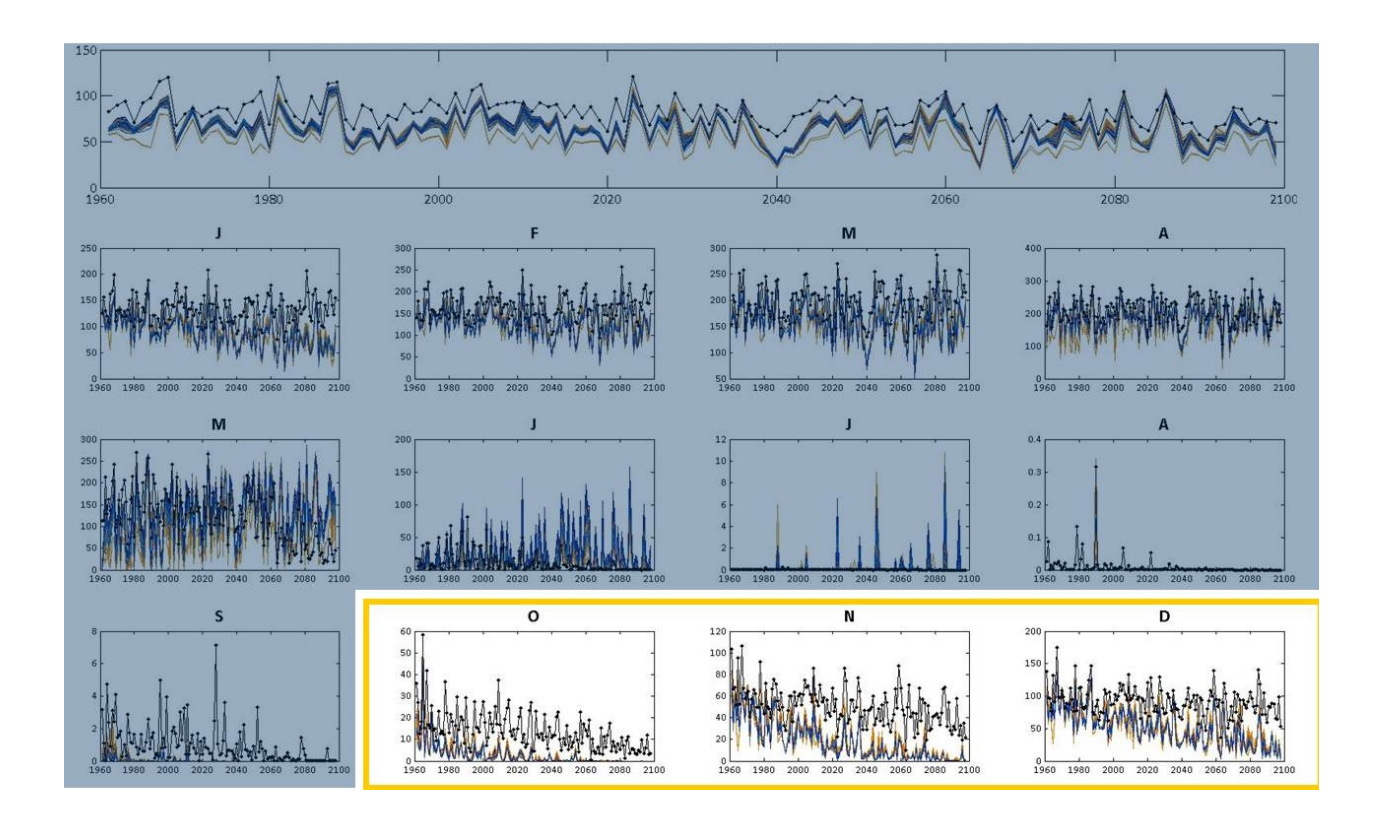


Figure G.21 Comparaison de la quantité de neige moyenne annuelle pour le modèle climatique, en noir, et pour les trois stratégies de calage appliquées au modèle hydrologique alimenté par la météo de la simulation ECHAM5#1, en orangé, bleu clair et bleu foncé, respectivement, pour les stratégies de calage traditionnel, traditionnel avec bilan ETR et parallèle.

cipitations sera reçu sous forme neigeuse. De la même façon, si la température est supérieure à 2° C, le modèle ne générera que de la pluie. L'intervalle entre les deux températures est géré par une combinaison linéaire pluie-neige proportionnelle aux températures. Le modèle climatique, quant à lui, effectue la séparation de phase de façon similaire, sur un intervalle [0 ; 6]. La zone grise entre 0 et 6° est cependant modélisée par une relation polynomiale de degré 6. Notons aussi que, comparativement au modèle hydrologique qui travaille avec des températures minimale et maximale journalières, le MRCC génère des données météorologiques à chaque pas de temps de quinze minutes. La séparation des phases est donc ajustée à chaque quart d'heure ce qui fait qu'il serait impossible de reproduire de façon exacte la méthode employée par un modèle climatique afin de l'imposer à un modèle hydrologique. Malgré tout, le sérieux d'un tel écart dans les précipitations neigeuses et des impacts que cela occasionne sur la capacité du modèle HSAMI à bien reproduire l'évapotranspiration ont motivé l'entreprise de démarches afin de corriger la situation.

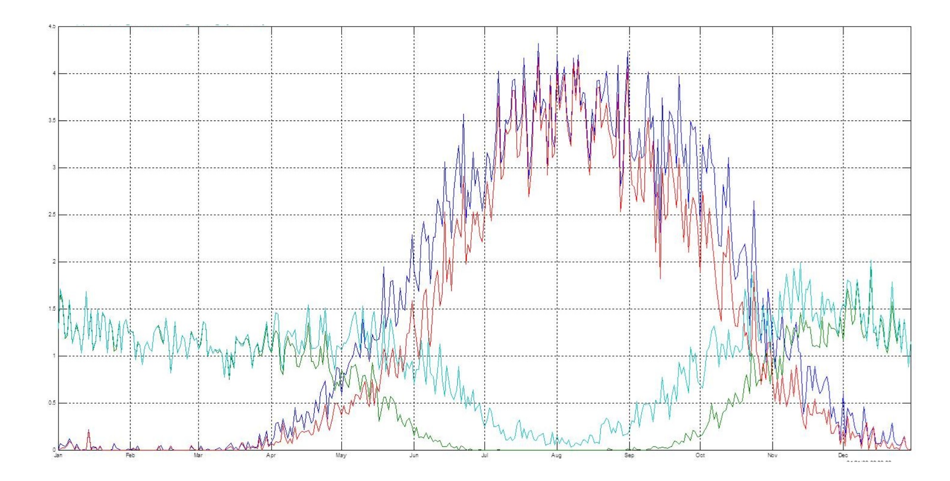


Figure G.22 Cycle annuel moyen où la pluie et la neige de la méthode [-2 ; 2] sont respectivement en bleu foncé et vert et où la pluie et la neige de la méthode [0 ; 6] sont représentées en rouge et cyan.

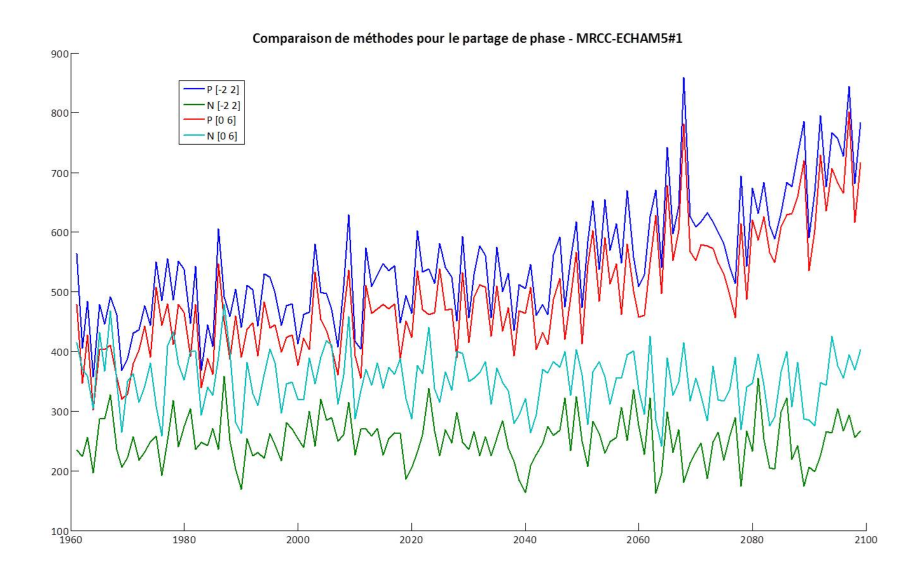
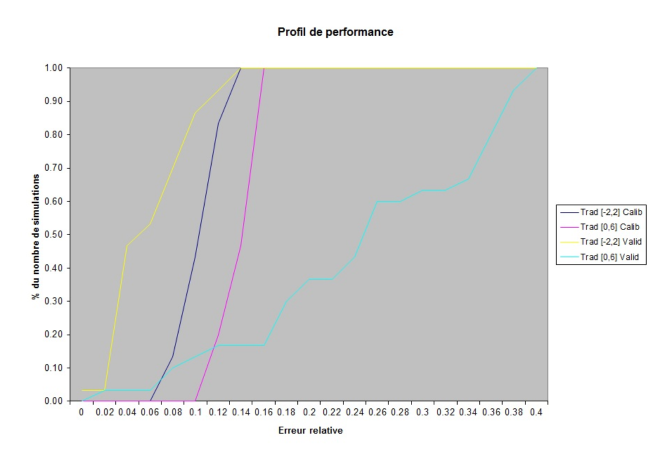


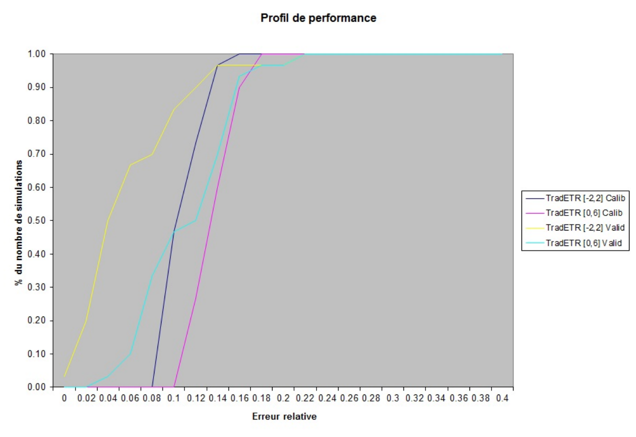
Figure G.23 Moyenne annuelle de la pluie (ici notée P) et de la neige (N) sur 140 ans.

Il a donc été convenu de mettre à profit l'information contenue dans les données météorologiques en se servant des températures minimale et maximale plutôt que de la moyenne journalière. Ainsi, la séparation des phases pouvait être effectuée toutes les douze heures. Cette stratégie a été couplée à la méthode [0 ; 6] du modèle climatique et a été testée sur le bassin Caniapiscau à l'aide des trois stratégies de calage. Malheureusement, non seulement le critère de Nash-Sutcliffe a-t-il été dégradé par rapport à la séparation de phases par la méthode [-2 ; 2], comme le montre la Figure G.24, mais les contraintes du calage en parallèle ont du être relâchées, signe que l'évapotranspiration était moins bien représentée. En effet,

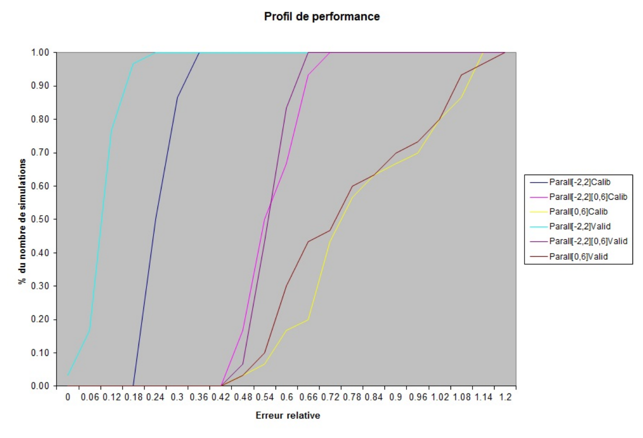
l'amplitude de l'évapogramme produit par cette méthode était inférieure à celle souhaitée, représentée par Qfs. Par contre, le problème du décalage a été résolu. Il a donc été supposé que la modélisation aux douze heures, plutôt qu'aux quinze minutes, était responsable puisqu'une quantité trop importante de neige a été reçue sur le bassin. Les quantités de neige et de pluie aux quinze minutes tels que modélisées par le MRCC sont conservées à Ouranos, mais n'étaient pas disponibles au moment d'écrire ces lignes. Cette piste a donc été temporairement abandonnée afin de conclure sur les stratégies de calage, mais sera définitivement reprise par M. Minville dans le futur.



(a) Performance des stratégies de séparation de phases avec un calage traditionnel



(b) Performance des stratégies de séparation de phases avec un calage traditionnel avec bilan sur l'ETR

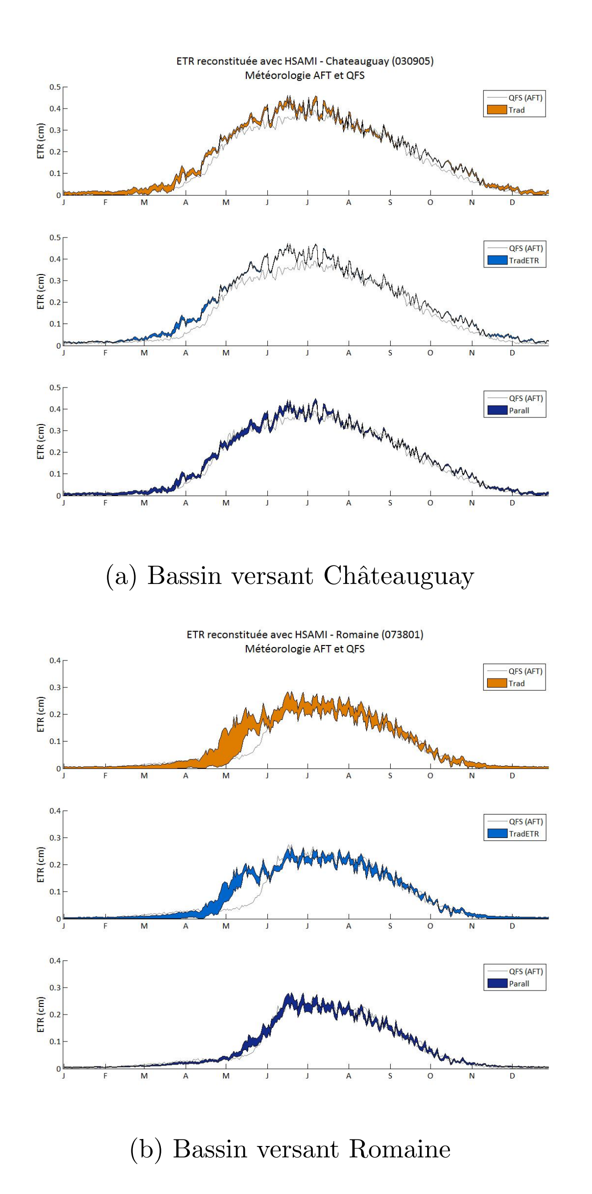


(c) Performance des stratégies de séparation de phases avec un calage parallèle

Figure G.24 Profils de performance pour les différentes méthodes de séparation de phases.

ANNEXE H

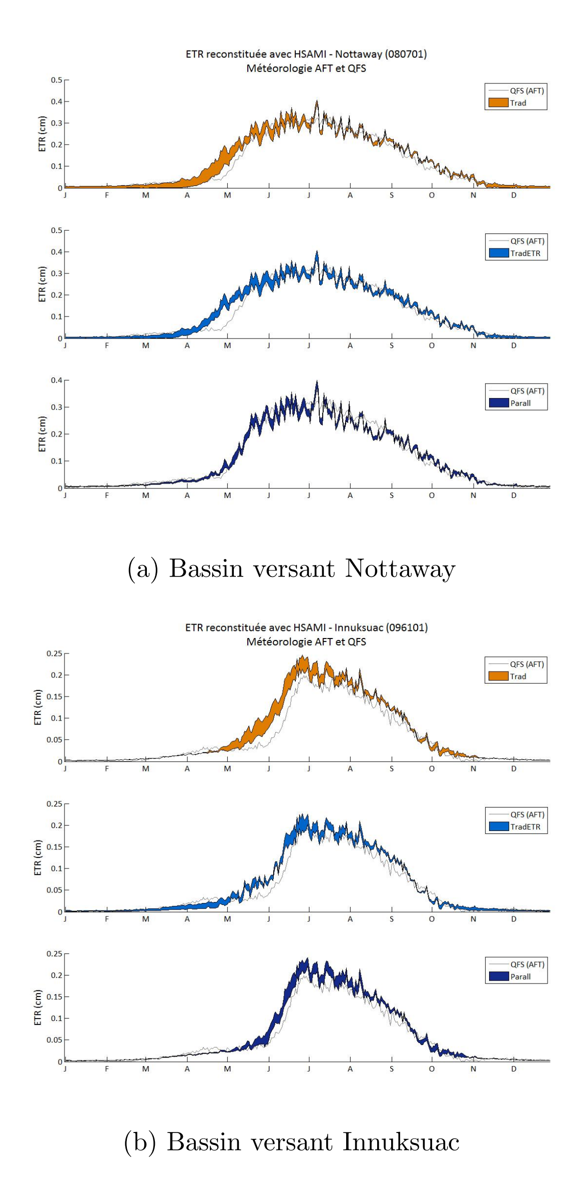
ÉVAPOGRAMMES DES TROIS STRATÉGIES DE CALAGE POUR LES SIX BASSINS



(a) Bassin versant Châteauguay

(b) Bassin versant Romaine

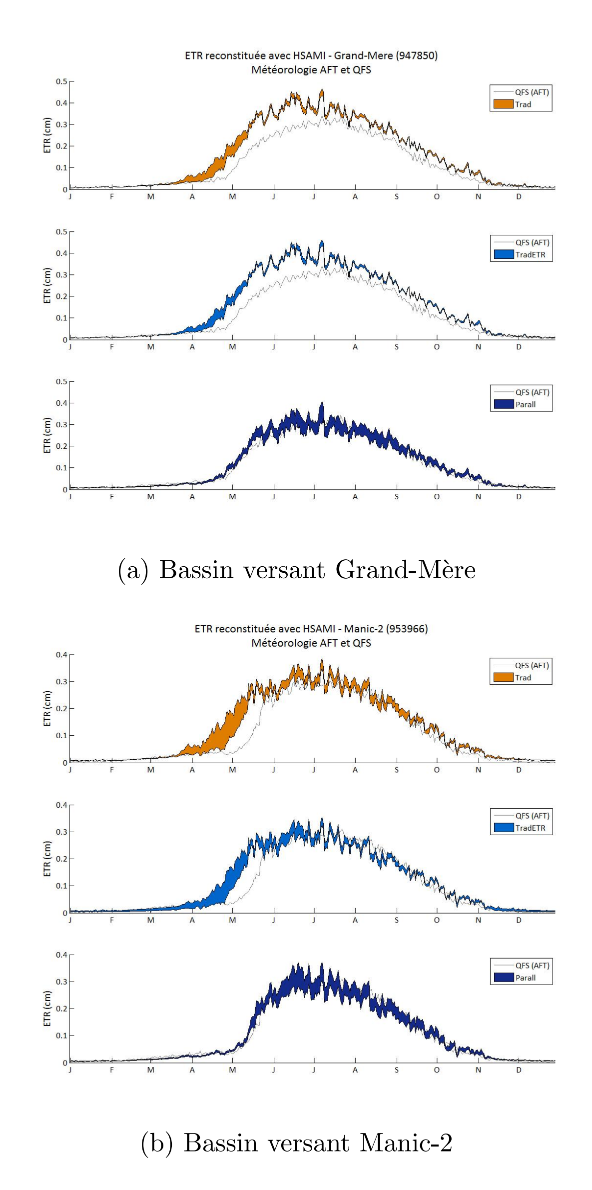
Figure H.1 Évapogrammes des trois stratégies de calage pour les bassins versants Châteauguay et Romaine.



(a) Bassin versant Nottaway

(b) Bassin versant Innuksuac

Figure H.2 Évapogrammes des trois stratégies de calage pour les bassins versants Nottaway et Innuksuac.



(a) Bassin versant Grand-Mère

(b) Bassin versant Manic-2

Figure H.3 Évapogrammes des trois stratégies de calage pour les bassins versants Grand-Mère et Manic-2.